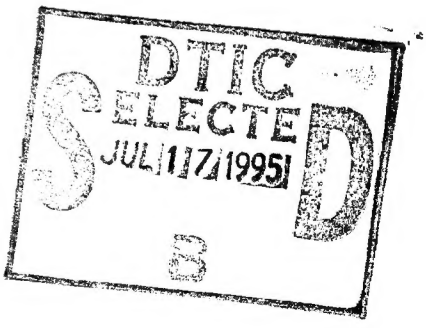
RL-TR-95-72
Final Technical Report
April 1995

FAST ALGORITHMS FOR BLIND SIGNAL SEPARATION AND CHANNEL EQUALIZATION

New Jersey Institute of Technology

Yeheskel Bar-Ness, Zoran Siveski, Alex Haimovich,
Nirwan Ansari, Raafat Kamel, and David W. Chen

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.



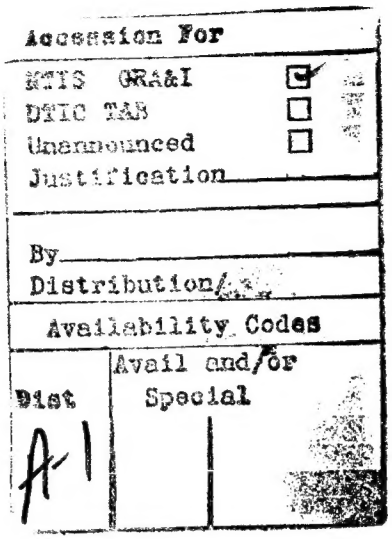
Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-95-72 has been reviewed and is approved for publication.

APPROVED:

RICHARD N. SMITH
Project Engineer



FOR THE COMMANDER:

HENRY J. BUSH
Deputy for Advanced Programs
Command, Control, and Communications Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (C3BA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1995	Final

4. TITLE AND SUBTITLE
FAST ALGORITHMS FOR BLIND SIGNAL SEPARATION AND CHANNEL EQUALIZATION

5. FUNDING NUMBERS
C - F30692-93-C-0118
PE - 62702F
PR - 4519
TA - 63
WU - P4

6. AUTHOR(S)
Yeheskel Bar-Ness, Zoran Siveski, Alex Haimovich, Nirwan Ansari, Raafat Kamel, David W. Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
New Jersey Institute of Technology
University Heights
Yeheskel Bar-Ness, Ph.D.
Newark NJ 07102-1982

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (C3BA)
525 Brooks Road
Griffiss AFB NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-95-72

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Richard N. Smith/C3BA/(315) 330-7436

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report covers a number of applications of the bootstrap algorithm. The algorithm is primarily a signal separator. As such, the algorithm has been applied to multi-user CDMA and that work is covered in this report. The convergence and stability of the bootstrap algorithm is also addressed. The report covers work on the application of the bootstrap algorithm to blind equalization of dispersive communication channels. A number of papers were published during the course of this effort and these papers are included as appendices to the report.

14. SUBJECT TERMS	15. NUMBER OF PAGES
communications, interference cancellation, adaptive, algorithms	180
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U/L

TABLE OF CONTENTS

Chapter **Page**

I. INTRODUCTION

Reported below are the results of a study carried out at the Center for Communications and Signal Processing Research at the New Jersey Institute of Technology between May 1, 1993 and July 31, 1994. This research is a continuation of previous research performed during similar periods in 1991/92 [1] and 1992/93 [2]. The aim of the research is to further widen the investigation of the bootstrapped algorithm to possible applications solving real system problems.

Historically, the idea of the algorithm as a way for canceling interferences was first proposed by the principle investigator in 1981 [3], and later it was used for canceling cross-polarization in satellite communication [4] and in the microwave terrestrial radio link [5 - 7].

The bootstrapped interference canceler is principally composed of two separate cancelers, each using the output of the other canceler as its reference (desired signal) input. In fact, such a structure performs as a "Signal Separator" rather than an interference canceler. Since for its operation there is no need for a reference signal, it is sometimes justifiably called a "Blind Separator."

Three different structures were proposed in [1]; namely, the Backward/Backward (BB), the Forward/Forward (FF) and the Forward/Backward (FB). They are depicted in Figures 1, 2, and 3. The different weights of these two-input/two-output separators can be controlled by minimizing the output power or minimizing the absolute value of the cross-correlation between the two outputs.

During the 1991/92 phase of the research [1], the steady state behavior of these separators was evaluated. The effect of additive noise on the performance of the separator was also examined. It was shown that extension of all three structures to multi-input/multi-output is possible. Therefore, it was expected that applications to multi-user Code Division Multiple

Access (CDMA) or neural networks is feasible. Only some preliminary work on the applicability of the algorithms was performed during this phase of the research and some results were presented. In particular, emphasis was put on the use of the three structures in handling a dually-polarized signal. Error probabilities of the bootstrapped cross-pol cancelers for M-ary QAM signals were examined. Their performance was compared with other kinds of cancelers known in the literature.

It was concluded at this phase of the research that the bootstrapped algorithms have many useful properties which make them excellent candidates for use as signal separators or interference cancelers when other algorithms have difficulties. In some cases, they clearly outperform other algorithms. In particular, it was demonstrated that the algorithm has the property of converging to its steady state, where signal separation occurs, much faster than other algorithms. Unlike other algorithms, the speed of convergence does not depend on the signals' power ratios and, hence, does not depend on the eigenvalue spread of the input correlation matrix.

The 1992/93 phase of the research [2] was aimed toward further analysis and evaluation of the performance of these algorithms. Inherently, the three structures exhibit different levels of complexity, particularly when used in very high frequencies, as in microwave communications. They also present different delay paths to signals and, hence, have different system bandwidths. Performance studies in this regard show the existing complexity-bandwidth trade off. The question of stability of the steady state was also examined. Two ways of real-time realization were proposed. One uses orthogonal perturbation sequences, while the other implements weight dithering with PN sequences.

Further work on delay controlled structures, which began in 1991/92, was continued in 1992/93. The work showed, in particular, that such a structure performs simultaneous spatial separation and direction-of-arrival estimation of wideband sources. With multi-input/multi-output applications, we obtained new results in blind separation of signals when

the environment is dispersive.

Some preliminary results were also obtained in applying these algorithms to separation of co-channel CDMA signals, as well as blind channel equalization.

Besides some issues of implementation, during the current phase of research (93/94) we concern ourselves with two separate directions of application: separation of multi-user CDMA signals and blind equalization of dispersive communication channels. Some neural network-type convergence studies were also performed. For each of these applications, system structures and block diagrams are suggested. Simulations are performed to prove applicability of the systems proposed. Some software modules were prepared to integrate the software in the Signal Processing Workstation environment.

Section (2), below, is a technical summary of the study performed during this research and its results. Detailed reports on which this summary is based are given in the appendices of the document. Each appendix covers a specific part of the research and is written in a way that can be read independently of the other parts. Section (3) contains the conclusions and recommendations for further studies.

II. TECHNICAL SUMMARY

2.1 Issues Concerning the Implementation of the Algorithms

The multi-input, multi-output structure of the bootstrap algorithm was considered in [1] (Appendices E and F). In Appendix G of [2] the performance of the backward/forward structure when it separates multi-signal composites in a multi-channel dispersive environment, is discussed. Further results for implementing these algorithms are given in Appendix A of this report.

In the first part, entitled, "Fast Decorrelation Algorithms for Signal Separation" the forward/forward structure of the bootstrapped algorithm, termed "decorrelation algorithm," is suggested for the separation of an unknown linear combination of source signals. The Steepest Descent (LMS) and Recursive Least Square (RLS) versions of the algorithm are considered. The results are compared with those of the conventional LMS algorithm, wherein, in contrast to the decorrelation algorithm, known references of the signal are assumed available. In particular, it is proved that the decorrelation algorithm exhibits a smaller eigenvalue spread and hence converges faster than the conventional LMS. In this study, the observation $x_m(t)$ is taken as a linear combination of the input sources, such that,

$$x_m(t) = \sum_i \sum_{n=1}^{N} A_{mn} b_n^{(i)} \sqrt{\xi_n} h_n(t) + \nu(t), \tag{1}$$

for $1 \leq m \leq N$. $b_n^{(i)} \in \{-1, 1\}$ is an information symbol bit of the n-th source at the i-th time interval. The information symbols are assumed to be statistically independent and equiprobable. $h_n(t)$ is the received pulse signature associated with the n-th source. The energy of $h_n(t)$ is normalized to 1, i.e., $\int_0^{T_s} h_n(t)dt = 1$ and ξ_n is the energy of the received signal from the n-th source, assumed orthogonal. Cross-correlations between the sequences are included in the matrix $\mathbf{A}$. $\nu(t)$ is (possibly correlated) additive Gaussian noise. It is assumed that $\mathbf{A}$ is strictly diagonal dominant, i.e., $\mid A_{ij} \mid > \sum \mid A_{ij} \mid$ with this assumption $\mathbf{A}$ is invertible. The coupling fixed parameters a_{mn} are not known. The received signals after

match filtering and sampling may be represented by the vectors:

$$\mathbf{x} = \mathbf{AEb} + \mathbf{v}, \tag{2}$$

where $\mathbf{A}$ is the mixture matrix, $\mathbf{E}$ is a diagonal matrix, $\text{diag}\mathbf{E} = \left[\sqrt{\xi_1}, \ldots, \sqrt{\xi_N}\right]^T$, and $\mathbf{b} = [b_1, \ldots, b_N]^T$ is the vector of information bits. The time superscript has been dropped since the decision on $b^{(i)}$ requires observation of signals during that interval only, thus $\mathbf{b}$ represents the vector of information bits at any time. The vector $\mathbf{v}$ consists of independent samples of zero-mean Gaussian noise with the covariance matrix $\text{E}\left[\mathbf{vv}^T\right] = \sigma_v^2\mathbf{G}$, where $\mathbf{G}$ is the known matrix of the noise cross-correlations between the outputs of the matched filter.

Clearly, the minimum square error criterion leads to the Wiener solution for the weight. The corresponding algorithm for controlling the weight is given by

$$\mathbf{w}_n(k) = (\mathbf{I} - \mu\mathbf{x}(k)\mathbf{x}^T(k))w_n(k-1) + \mu\hat{b_n}(k)\mathbf{x}(k), \tag{3}$$

where $\hat{b_n}(k) = \text{sgn}y_n(k)$ is the decision at the n-th output y_k.

It was also shown that the criterion for output decorrelation can be obtained by minimizing the cost $J = \mathbf{w}_n^T R_{\hat{b_x}}\mathbf{w}_n$ where $R_{\hat{b_x}} = \text{E}[\hat{b}\mathbf{x}^T]$ under the constraint that $\mathbf{w}_{nn} = 1$. The corresponding algorithm for controlling the weight is given by

$$\mathbf{w}_n(k) = \left[\mathbf{I} - \hat{B}(k)\mathbf{x}^T(k)\right]\mathbf{w}_n(k-1), \tag{4}$$

where $\hat{B} = (I - \mathbf{u}_n\mathbf{u}_n^T)\hat{b}$ and u_n is a unit vector with the n-th element equal to 1. From comparing (3) and (4) it was possible to deduce that the eigenvalue spread of the first algorithm is larger than that of the second. Simulation results of learning curves are given in Fig. 1 and Fig. 2 of Appendix A, part 1. Shown in Fig. 3 and Fig. 4 is an estimate of the convergence region using a γ factor defined by $\gamma = 1 - p_{cf}/p_{ci}$, where p_{cf} and p_{ci} are the final and initial probabilities of error.

In conclusion, this study showed that the LMS decorrelator algorithm is faster than the LMS error algorithm due to a smaller eigenvalue spread. We introduced a new RLS-type

algorithm for decorrelation. The RLS decorrelator was shown to be faster than the LMS-type algorithms and of comparable speed with the conventional RLS error. Regions of convergence for the LMS decorrelator and the RLS algorithms were shown to be wider than those of the LMS error algorithm.

In the second part of Appendix A entitled "Adaptive Detectors for Multi-Channel Signals in Code Division Multiple Access System," both the feedforward and feedback structure are considered for implementation in CDMA systems. However, an assumption was used that at the output of the bank of matched filters, regularly implemented in this application, the cross-correlation matrix $\mathbf{A} = \mathrm{A}_{ij}$ is unknown. This fact might be cause by the unknown channel effect. Nevertheless this fact lead to a problem model similar to that of the previous part of the appendix. Four adaptive detection criteria are considered: two each for the forward structure using MSE and decorrelating. The other implements feedback structures, again with MSE and decorrelating. The optimal weights for all these algorithms and structures are found. Error probability is used as a performance measure for comparison. It is shown that the feedback minimum square error has the best performance in terms of probability of error. This might be due to the fact that in the MSE we assume the availability of a reference as well as a noise-free case.

2.2 Application to Signal Separation of Multi-User CDMA Signals

In this section we concentrate on signal separation of a multi-user detector for CDMA a system. For such a system, the receiver signal is given by

$$r(t) = \sum_{k=1}^{k} \sum_{i} b_k(i)\sqrt{a_k}s_k(t - iT - \tau_k) + n(t), \tag{5}$$

where $b_k(i) \in \{-1, +1\}$ is the k-th user data bit in the i-th time interval, $s_k(t)$, k=1, ...,K is a unit energy signature waveform of duration T assigned for each of k different users. $n(t)$ is a zero-mean white Gaussian noise with the two-sided power spectral density $N_0/2$, and a_k and τ_k are the received energy and relative delay of user k. The energy a_k is assumed

unknown, particularly due to fading.

If $r(t)$ is the signal received at the main station (uplink) then τ_k might be different for different users. Though it is unknown, sometimes it is assumed to be estimated by a separate process. If, on the other hand, $r(t)$ is the received signal at a single-user location (downlink), then τ_k=0 for every k and the signal is considered synchronous.

2.2.1 Adaptive Separation of the Synchronous Multi-User CDMA Signal

The model for this signal separator is given in Figure 4. Some variation to this scheme will be discussed in the sequel. Since τ_k=0, k=1, ..., K then one can show that vector $\mathbf{x}$, after the matched filters, is given by;

$$x(i) = \mathbf{P}\mathbf{A}b(i) + n(i), \tag{6}$$

where $\mathbf{x}$=$[x_1, \ldots, x_K]$, $\mathbf{b}$=$[b_1, \ldots, b_K]$, $\mathbf{A}$ = diag$(\sqrt{a_1}, \ldots, \sqrt{a_K})$ and $\mathbf{n}$=$[n_1, \ldots, n_K]$. The $kj-$th element of the symmetric cross-correlation matrix $\mathbf{P}$ is defined as

$$\begin{aligned} P_{kj} &= \int_0^T s_k(t)s_j(t)dt \quad k,j \in (1,2,...K) \qquad (7) \\ n_j &= \int_0^T s_j(t)n(t)dt \\ P_{kk} &= 1. \qquad (8) \end{aligned}$$

The covariance matrix of the zero Gaussian noise vector $\mathbf{n}$ is given by $(N_0/2)\mathbf{P}$.

Multiplication by $\mathbf{P}^{-1}$ (assumed known) leaves vector $\mathbf{z}$ having uncorrelated signals

$$\mathbf{z} = \mathbf{A}\mathbf{b} + \xi, \tag{9}$$

where $\xi = \mathbf{P}^{-1}\mathbf{n}$.

At the canceler output we generate an output y_k by multiplying a function of z_i by weights w_{jk} and subtract the result from x_k. Therefore, in vector notation

$$\mathbf{Y} = \mathbf{X} - \mathbf{W}^T f(\mathbf{z}), \tag{10}$$

where

$$\mathbf{W} = \begin{bmatrix} 0 & w_{12} & \dots & w_{1K} \\ w_{21} & 0 & \dots & w_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ w_{K1} & w_{K2} & \dots & w_{KK} \end{bmatrix}.$$

The steepest descent algorithm is used to minimize the output signal power $E(y_k^2)$:

$$\mathbf{w}_k \leftarrow \mathbf{w}_k + \mu \frac{\partial}{\partial \mathbf{w}_k} E(y_k^2), \tag{11}$$

where $\mathbf{w}_k$ is the k-th column vector of $\mathbf{W}$.

Note that since z_j, j=1, ... K (j$\neq$ k) contains, besides the data b_j, a noise component ξ_j, the error performance at the output will depend strongly on SNR's at z_j. This is exactly what is known as power inversion behavior inherent in LMS and minimum pole cancelers. To improve performance sgnz_j is used instead of z_j. Particularly when SNR at z_j for all j are high, the canceler will perform at its best.

In Appendix B, Part I, we analyze the separator depicted in Figure 4; we derive the optimal weights and then use them to calculate the probability of P_{ek} at the k-th output as a function of cross-correlation vector ρ_k is a column vector obtained from $\mathbf{P}$ by deleting the element ρ_{kk}.

For the two-user case, we show in Figure 5 the probability of error of user 1 as a function of $SNR_2 - SNR_1$ with ρ=0.7 and SNR_1=8dB. For comparison the probability of error at the output of decorrelator is also shown. As expected for high SNR_2 the estimate of b_2 from z_2 is almost perfect and the canceler performs at its best (ideal separator). For a very low $SNR_2 - SNR_1$ the error at y_1 is better due to low interference at x, the output of the matched filter.

Figure 6 depicts the error probabilities when all (K-1) users except user 1 maintain the same SNR_i. Gold codes are implemented in this figure and K was taken to be 2 to 5.

2.2.2 Adaptive Separator of a Synchronous Multi-User CDMA Signal with Soft Decision

As stated in the previous section, the signal separator of Figure 5 performs better at high $SNR_i - SNR_1$ and at low $SNR_i - SNR_1$. Using a hard decision (Signum function) at low SNR_i is not advised, due to a higher error in estimating the interference. This suggests using a soft decision instead. With this we have

$$y_k = x_k - \mathbf{w}_k^T \mathbf{h_k}, \tag{12}$$

where

$$h_{lk} = \begin{cases} z_l/\lambda_{lk} & \mid z_i \mid < \tau_{lk} \\ sgn(z_l) & \text{otherwise.} \end{cases} \tag{13}$$

The value of λ_{lk} is determined heuristically from observed values of decorrelator outputs as

$$\lambda_{lk} = \rho_{lk} \frac{(E\{\mid z_k \mid\})^2}{E\{\mid z_l \mid\}}. \tag{14}$$

In Appendix B, Part II we analyze this separator, we derive the optimal weights and use them to calculate the probability of ρ_{ek} at the k-th output. In Figure 7 we show the probability of error of user 1 as a function of $SNR_2 - SNR_1$ with $\rho = 0.7$, $SNR_1 = 8$dB for the two-user case. More results are also shown in this appendix. For comparison we also add to this figure the performance of the decorrelation and the canceler with a hard limiter. It is easy to recognize that using a soft limiter improves performance for almost all SNR_2. Particularly noticeable is the improvement at a low SNR_2 value, where the hard limiter results in poor performance.

2.2.3 Adaptive Separator of Synchronous Multi-User CDMA Signals Using the Bootstrapped Decorrelating Algorithm

Here, instead of using an algorithm that minimizes the output signal power $E(y_k^2)$ (see equation 11), we use a steepest descent algorithm, which simultaneously reduces the absolute value of the correlation between the output y_k and the decision on all other outputs, *i.e.*,

$$\mathbf{w}_k \leftarrow \mathbf{w}_k - \mu E\{y_n sgn(\mathbf{y}_k)\}. \tag{15}$$

This is what we referred to in the previous report as the forward/forward bootstrapped algorithm. In Figure 8 we depict the block diagram of this separator. For obtaining the optimal weight one has to solve a system of nonlinear equations. However, if we assume the signal-to-noise ratio at the output is such that the main contribution to the output error is due to multiuser interference, then we can use the following approximation:

$$E(y_k sgny_k) \simeq E(y_k \mathbf{b}_k)(\mathbf{I} - 2P_{ek}), \tag{16}$$

where P_{ek} is a diagonal $(K-1) \times (K-1)$ matrix whose elements are P_{ej}, j$\neq$ k = 1, 2, ... $K-1$, $K+1, K$. This approximation changes the nonlinear system of the equation to a tractable linear system. In Appendix B, Part III we present analysis of this separator, we derive the optimal weights, then we calculate the probability of error. In Figure 9 we show the results of this separator for the two-user case.

2.2.4 One-Shot Adaptive Separator of an Asynchronous Multi-User CDMA Signal

Different from the synchronous separator, some or all τ_k in (5) are non-zero. However, we will assume them to be known. Without loss of generality we concentrate on bit 0 of user 1, and assume that $0 = \tau_1 \leq \tau_2 \leq \ldots \leq \tau_k < T$. The sampled output of the matched filter for

user 1 is then

$$x_1(0) = \sqrt{a_1}b_1(0) + \sum_{k=1}^{K} \sqrt{a_k}\,[\rho_{k1}b_k(-1) + \rho_{1k}b_k(0)] + n_1(0). \tag{17}$$

The normalized partial cross-correlations ρ_{k1} and ρ_{1h} for k = 2 ,... , K are

$$\rho_{k1} = \int_0^T s_1(t)s_k(t+T-\tau_k)dt \tag{18}$$

$$\rho_{1k} = \int_0^T s_1(t)s_k(t-\tau_k)dt, \tag{19}$$

and

$$n_1(0) = \int_0^T n(t)s_1(t)dt$$

is a zero-mean Gaussian random variable with a variance of $N_0/2$. In matrix notation:

$$x_1(0) = \sqrt{a_1}b_1(0) + \boldsymbol{\rho}_1^T\mathbf{A}\mathbf{b}_1(0) + \mathbf{n}_1(0), \tag{20}$$

where $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_1 \end{bmatrix}$, $\mathbf{A}_1 = \text{diag}\,[\sqrt{a_2}, \ldots \sqrt{a_k}]$, $\text{b}_2(0) = [\text{b}_2(\text{-}1), \ldots, \text{b}_k(\text{-}1), \text{b}_2(0), \ldots \text{b}_K(0)]$ and $\boldsymbol{\rho}_1 = [\,\rho_{21}, \ldots \rho_{21}, \rho_{12}, \ldots, \rho_{1k}\,]$. The separator forms an estimate of the multiuser interfere as the weighted vector of tentative decisions on symbols that interfere with $b_1(0)$ directly, and subtracts it from the matched filter output.

Except for the large dimension of the weighted vector, the canceler section is the same as in the synchronous case. Obviously, depending on τ_k, the vector $\boldsymbol{\rho}_k$ will change the relative powers in the current and previous bits of the interferers and so the performance of the canceler.

In Appendix B, Part IV we present, some analysis and results, primarily with two users of this separator. Figure 10 shows the probability of error of use 1 versus SNR_1 for the two asynchronous users. The relative energy is defined as $\int_{\tau_2}^{T} s_1^2(t)dt$. For comparison we add to this figure the conventional single-user detector and the single-user bound when interference is absent.

2.2.5 Adaptive Bootstrap Decorrelator for Multi-User CDMA in an Asynchronous Unknown Channel

In any of the previously discussed adaptive separators, the principle idea is that of using a bank of matched filters at the input of the receiver. If the users are synchronous and the codes are known, then besides noise, the outputs of the filters contain a known mix of the different users' bits. This mixture depends on the cross-correlation of the codes. Linear transformation on these outputs with the inverse of the cross-correlation matrix will generate outputs that contain a separate signal (decorrelator). The separation is inadequate because of the high noise contamination. A tentative decision followed by an adaptive canceler did the job of reducing the interference at the output of the matched filter without having high noise.

If on the other hand the users are asynchronous but the relative delays and the codes are known, then the one-shot approach will again generate a known mixture at the outputs of the bank of the matched filter, which corresponds to each user bit 0 (for example) and bit (0) and bit (-1) of other users. Again, linear transformation by the inverse of this generalized cross-correlation matrix will result in an output with uncorrelated signals. Then, as before, a tentative decision followed by an adaptive canceler will do the job.

For the asynchronous unknown relative delay τ_k, the mixture at the outputs of the matched filters are unknown. This is also the case if τ_k's are estimated by some process, then there will be some estimation error. For these cases we suggest using the bootstrapped algorithm for decorrelating the outputs of the matched filters. Such a decorrelation is proposed in Appendix B, Part V.

The principle operation of this decorrelator (see Figure 11) is weighting the outputs of the matched filter by a matrix $\mathbf{W}$, the same as in (10), while the next and previous bit by upper- and lower-triangular parts of the other matrix $\tilde{\mathbf{W}}$, that is, a total of 2(K-1) weights for each user output (total of 2K(K-1)) weights. The adaptive algorithm then controls the weight

matrices $\mathbf{W}$ and $\tilde{\mathbf{W}}$ to reduce the absolute value of the correlation between the decorrelation output $y_m(i)$ and all other outputs of the decorrelator (after decision) at time τ, i-1 and i+1 respectively. The weights $\tilde{w}_{mk}$ and $\tilde{w_{mk}}$ for 1 ¡ K, m $\leq$ K, a $\neq$ m are controlled.

$$\begin{aligned} w_{mk} &\leftarrow w_{mk} + \mu E\left(y_m(i) sgn\left(y_k(i)\right)\right) \\ \tilde{w}_{mk} &\leftarrow \tilde{w}_{mk} + \mu E\left(y_m(i) sgn\left(y_k(i-1)\right)\right) \quad k > m \\ \tilde{w}_{mk} &\leftarrow \tilde{w}_{mk} + \mu E\left(y_m(i) sgn\left(y_k(i+1)\right)\right) \quad k < m. \end{aligned} \tag{21}$$

Note that for the two-user case K=2, there are 4 weights, w_{12}, w_{21}, $\tilde{w}_{12}$ and $\tilde{w}_{21}$.

In Appendix B, Part V we present a more detailed analysis of this decorrelator. Particularly for the two-user case, we found the optimal weight and use it to derive the error probability of user 1 at the output of the decorrelator. This appendix also contains some simulations and analytical results, and compares them to the results obtained with the one-shot, when τ_k is known.

2.2.6 Convergence and Stability Analysis of the Adaptive Separators

In Appendix B, Part VI we consider the question of convergence of the weight control algorithm to the optimal value. Although this is done for the synchronous case when the algorithm minimizes the output power, the derivation can be extended to the decorrelation algorithm with appropriate linearization. Extension includes the possibility of the one-shot asynchronous case. This study concludes with a condition on the learning rate μ.

In summary, we presented in this section two approaches for adaptive separation of multi-user CDMA signals, and we show their performance in terms of error probability. All these separators are blind cancelers, since they only use the received signal without the help of training signals. Some of the algorithms used a fast algorithm; others are minimum power algorithms whose performance is enhanced with tentative predecisions.

2.3 Blind Equalization of Dispersive Communication Channels

In digital communication the need for blind equalization arises when the channel is unknown to the receiver, which attempts to estimate a sequence of transmitted data without resorting to the use of a training sequence. The main problem of blind equalization is that of finding an appropriate cost function (or equivalent error function) that reflects the amount of intersymbol interference (ISI) introduced by the channel, and which does not involve the transmitted symbol. Optimization of the cost function should lead to minimization of the ISI. Many researchers dealt with blind equalization and the ill convergence problem, which was treated with modifications suitable for the case, but with extra complexity.

The main emphasis of their research is on the linear equalizer structure. Another very important structure in digital communication, the decision feedback equalizer, receives less attention.

As part of our current research effort, we suggest, implement and study of the bootstrapped decorrelation algorithm in conjunction with decision feedback equalization. Another topic of blind equalization considered in this phase, although not related to the decorrelation algorithm, is the "Anchored Blind Equalizer using the Constant Modulus Algorithm" for improving performance.

These two topics belong to a larger class of symbol-by-symbol equalization. The second topic involves maximum likelihood sequence estimation (MLSE), where the Viterbi algorithm (VA) is used to search a trellis for the desired sequence. The blind maximum likelihood (ML) sequence estimation problem is discussed in section 2.3.3 and the corresponding part of Appendix C. Here the linear dispersive channel model is assumed unknown, but it has quantized parameters. A channel trellis and a data trellis are defined to search for the ML channel and data estimates using the Viterbi algorithm. The approach provides a good performance/complexity tradeoff.

In contrast to symbol-by-symbol estimation, maximum likelihood sequence estimation has a lower error rate at the expense of complexity. The complexity of the VA grows exponentially with the length of the channel response. When the impulse response becomes larger, the VA becomes impractical, and methods for complexity reduction are needed. In section 2.3.4 and in the corresponding part of Appendix C, we present a new method of complexity reduction of such an equalizer.

2.3.1. Blind Decision Feedback Equalization Based on Decorrelation

The structure of such an equalizer is shown in Figure 12. The cascade of the transmit channel and receiver filters is modeled as an FIR filter with impulse response

$$h(n) = 1 + \sum_{i=1}^{N} h_n \delta(n-i), \tag{22}$$

where $\delta(\cdot)$ is the kronecker delta. It is also assumed that the input I_k is a binary white sequence with a zero mean. The output of the channel is thus given by

$$x_k = I_k + \sum_{i=1}^{N} h_i I_{k-i}. \tag{23}$$

Clearly the post-cursors $(h_1, h_2, ..., h_N)$ introduce ISI. The estimated data $\hat{\mathbf{A}}_k$ is produced by passing A_k through a slicer. The input to the slicer of the decision feedback equalizer A_k is given by

$$\begin{aligned} A_k &= X_k - \hat{\mathbf{A}}'_{k-1}\mathbf{W} \\ &= I_k + \mathbf{I}'_{k-1}\mathbf{H} - \hat{\mathbf{A}}'_{k-1}\mathbf{W}, \end{aligned} \tag{24}$$

where $\hat{\mathbf{A}}_{k-1}$ is the vector of the past N decisions $\hat{\mathbf{A}}'_{k-1}$=$[\hat{A}_{k-1}, \hat{A}_{k-1}, \ldots, \hat{A}_{k-N}]$ (the prime stands for transpose) and $\mathbf{I}_{k-1}$ is the vector of past transmitted information bits $\mathbf{I}'_{k-1}$ = [$I_{k-1}, I_{k-2}, \ldots, I_{k-N}$], where $I_{k-i} \in$ { -1, 1 } and $P\{I_{k-i} = 1\} = \mathrm{P}\{I_{k-1} = -1\} = 1/2$. $\mathbf{W}$ and $\mathbf{H}$ are the equalizer and channel parameter vectors, respectively; $\mathbf{W}$' = [$w_1, w_2, \ldots, w_N$] and $\mathbf{H}$' = $[h_1, h_2, \ldots, h_N]$.

A decorrelation algorithm which uses $\overline{A_k A_{k-i}}$ as an error function adaptively controls the equalizer weight according to

$$w_i^{k+1} = w_i^k + \mu \overline{A_k A_{k-i}} \quad i = 1, 2, ..., N. \tag{25}$$

In a practical implementation one would replace the expectation by the current realization, leading to a stochastic difference equation.

For ideal ISI cancellation, the slicer's input $A_k = I_k$, and therefore sequence $\{ A_k \}$ will be decorrelated, *i.e.*, $\overline{A_k A_{k-i}}$ =0 for n=0. In other words, decorrelation is a necessary condition for ideal cancellation of ISI. In order to be able to use the decorrelation of the slicer's input as a criterion for controlling the feedback weight vector **W**, one must prove that decorrelation is also sufficient for cancelling ISI. This is done in detail in Appendix A, Part I. Simulation performed on different channel models showed convergence to the desired values of weights, and hence ISI cancellation. It was also shown that the decorrelation algorithm converges to the correct weight irrespective of the initial condition.

Upper bounds and lower bounds on the probability of bit error for the AWGN case with a zero mean and a variance of σ^2, was found. A channel whose transfer function is given by $H(z^{-1})=1 = 0.8\ z^{-1} + 0.6\ z^{-2} + 0.4\ z^{-3}$, is depicted in Figure 13, these bounds together with results of simulation of this channel and the no ISI probability of error are shown for comparison. Due to the simplicity of the decorrelation algorithm, one could derive a Kalman-type decorrelation blind equalizer, which increases the convergence substantially. This is included in Appendix C, Part I.

2.3.2 Anchored Blind Equalization Using the Constant Modulus Algorithm

As stated earlier, most algorithms for the blind equalizer in the literature use a non-convex cost function that possesses local minima to which the equalizer may converge. Some of these equilibria may be undesirable, *i.e.*, they will not be able to remove ISI. For these algorithms,

equalizer initialization becomes an important issue. Verdu showed that anchoring (setting the first coefficient to one) will result in global convergence. Verdu used the minimum energy (ME) as a cost function. An important algorithm developed by Triechler is the constant modulus algorithm (CMA) which, like the others, suffers from ill-convergence. In Appendix C, Part II, we suggest an anchored constant modulus algorithm (ACMA) for the linear and decision feedback blind equalizers (see Figure 14, for example).

Detailed analysis of this algorithm is given in this appendix for equalizing auto-regressive channels and moving average-type channels. It is shown analytically and through simulation that the algorithm converges successfully if the unknown channel gain exceeds a certain value $(1/\sqrt{3})$. The controlled weight will fail to converge to the desired value if the channel gain drops below this value. This problem can be minimized if we introduce a gain in front of the equalizer. In comparing the ACMA to that of AMEA we notice that the former converges faster than the latter. With AMEA we notice that the former converges faster than the latter (see Figure 16). The AMEA, however is showing no ill convergence for a small g, but like the ACMA rate of convergence, it decreases with decreasing g.

2.3.3 Blind Maximum Likelihood Sequence Estimation

For sequence estimation of digital data in the presence of intersymbol interference, both the data and the channel might be considered unknown. To facilitate the proposed approach of this research, we assumed a quantized channel. This is justified in practice since a finite precision processor is used to implement the algorithm. Two trellises are developed; one for the channel and the other for the data. The VA is used to search for the ML channel and data sequence estimates. The output of one trellis is fed into the metric calculator of the other. The resulting scheme offers a considerable reduction in computational complexity compared with other methods available in the literature, and prevails with regard to the complexity/performance tradeoff.

The sampled output of the channel, r_k at instant k, is given by

$$r_k = \Gamma \mathbf{I}_k' \mathbf{h} + n_k, \tag{26}$$

$\mathbf{I}_k = [I_k, I_{k-1}, \ldots, I_{k-l}]'$, $\mathbf{h} = [\ h_0, h_1, \ldots, h_L]'$, and $\{h_i\ \}_{i=0}^{L}$ is the sampled impulse response of the cascaded transmit channel and receiver filters are assumed to be slowly time varying. $\{I_{k-i}\}$ is the sequence of transmitted symbols; assumed to be iid random variables whose values are $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$ with equal probability, and $\{n_k\}$ is an additive white Gaussian noise sequence.

For the blind equalization problem at hand, we consider the conditional probability of the received sequence $\mathbf{r}$ given the transmitted sequence and the channel impulse response. With all channel realizations equally probable, the joint ML estimate for the data sequence $\mathbf{I}_{ML}$ and the channel parameter $\mathbf{h}_{ML}$ is given by

$$(\mathbf{I}_{ML}, \mathbf{h}_{ML}) = \arg \max_{\mathbf{I},\mathbf{h}} f_{\mathbf{r}|\mathbf{I},\mathbf{h}}(\mathbf{r}|\mathbf{I}, \mathbf{h}), \tag{27}$$

wherein the maximization is carried over all possible $\mathbf{h}$ and $\mathbf{I}$. $f_{\mathbf{r}|\mathbf{I},\mathbf{h}}(\cdot \mid \cdot, \cdot)$ is the conditional probability density function (pdf) of the received sequence $\mathbf{r}$. Such a problem is not trivial since $\hat{\mathbf{h}}$ is continuous and $\mathbf{I}$ is discrete.

The channel parameters are approximated by discrete values from the infinite alphabet $\{0,\ \pm c, \pm 2c, \ldots\ \}$, where c can be chosen as arbitrarily small. The corresponding channel trellis will have an infinite number of states. However, since the channel vector $\mathbf{h}$ does not vary over each signal interval, we need not consider all its possible levels at a given instant. We proposed a simple assignment scheme for the channel trellis. The next channel estimate $\mathbf{h}^{i+1}$, is given by

$$\begin{aligned} \mathbf{h}^{i+1} &= \mathbf{h}^{i} \quad \text{for state 0, and} \\ \mathbf{h}^{i+1} &= \mathbf{h}^{i} \pm c\mathbf{1}_n \quad \text{for state } n = 1, 2, \ldots, \end{aligned}$$

where $\mathbf{1}_n$ is a vector of length L, whose elements are either 0's of 1's.

A smaller number of states can be used if the vector $\mathbf{1}_n$ is restricted to be all zero, except for the element at the $n-th$ location to unity. This assignment results in $L+2$ states. There are other schemes, up to 2^{L+2} states, but the above assignment will result in a simple trellis.

In summary, the algorithm would proceed as follows:

1. Start with initial channel estimate, $\mathbf{h}^0$,

2. Use the VA to solve for $\hat{\mathbf{I}}_{\mathbf{ML}} = \arg\max_{\mathbf{I}} f(\mathbf{r}|\mathbf{I}, \hat{\mathbf{h}}_{ML})$,

3. Use the VA to solve for $\hat{\mathbf{h}}_{\mathbf{ML}} = \arg\max_{\mathbf{h}} f(\mathbf{r}|\hat{\mathbf{I}}_{ML}, \mathbf{h}_{\mathbf{ML}})$,

4. Iterate 2 and 3.

The algorithm achieves the ML estimate of the channel by incrementing or decrementing the previous estimate. In each transition only one channel parameter is changed. To improve the speed of convergence one can include more states in the channel trellis, which allows for changing two or more parameters at a time, at the expense of complexity. Thus, one can compromise the rate of convergence for complexity. The step parameter c affects the performance. Choosing a smaller c will reduce the rate of convergence, but it will improve the error rate.

The algorithm described above was used to equalize the channel (assumed unknown) whose sampled impulse response is given by

$$h(n) = 0.405\delta(n) + 0.817\delta(n-1) + 0.407\delta(n-2).$$

For simplicity, binary data is assumed. The channel and data trellises, each having four states, are shown in Figure 15. Clearly, using smaller c results in higher complexity but better performance.

2.3.4 Reduced Complexity Sequence Estimation Using State Partitioning

As was stated earlier, the maximum likelihood sequence estimator has a lower error rate at the expense of complexity. The complexity of the VA grows very fast with the length of the channel and or the size of the constellation. Many researchers have been considering the problem of reducing state complexity and came out with different proposals. Some of these are useful for systems utilizing a larger signal constellation, while others are suitable for channels with a long impulse response. We proposed a state reduction method which is based on a state partitioning technique, which can handle both a large constellation and long impulse response channels. Moreover, it is richer in the possible number of reduced states (not only power of M, where M is the alphabet size). Therefore, it enables a more flexible complexity/performance tradeoff.

The discrete time channel model considered here is given in Figure 16. This model arises in a quadrature amplitude modulation (QAM) system at the output of a sampled, whitened matched filter. The channel $h(D)$ is modeled as a finite response filter (FIR), and $n(D)$ is a white Gaussian noise source with a zero mean and variance of σ^2. The data sequence $a(D)$ consists of symbols a_k, which are independent and identically distributed. We will assume binary transmission here.

The data symbols a_k take values of ± 1 with equal probability. Referring to Figure 17, the output $y(D)$ is given by $y(D) = a(D)h(D) + n(D)$, where $h(D)$, given by $h(D) = h_0 + h_1 D + \ldots + h_m D^m$, defines the channel impulse response, whose degree is determined by the channel memory. The state of the system, $s(k)$, at time k is given by $(a_{k-1}, a_{k-2}, \cdot, a_{k-m})$. Denote the set of all 2^m possible states by $\mathcal{E}$; then

$$\mathcal{E} = \{s_i : s_i \text{ is a state of the system}, i = 0, 1, \ldots, 2^m - 1\}.$$

In the proposed method, the set $\mathcal{E}$ is partitioned into N, S_i, partitions, where N is $2 \leq N \leq 2^m$, such that

1. $\bigcup_{i=0}^{N-1} S_i = \mathcal{E}(k)$,

2. $S_i \cap S_j = 0$; the empty set for $i \neq j$ and $0 \leq i, j \leq N-1$,

3. The partitions S_i are chosen such that for all $s_n(k) \in S_i$, the corresponding next state $s_n(k+1)$ must belong to the same partition.

In particular, the third condition contains the partition in a way that enables a trellis to be defined. A procedure was devised in which the partitions satisfy the third rule. Detailed information is given in Appendix C, Part IV.

In this appendix we show how the method can be applied to obtain trellises with 2^i states as well as trellises with the number of states not 2^i. Bounds on the error probability of these cases are also found. Applications to non-binary (M-PAM, M-QAM and MPSK) modulation are also discussed. Simulations were performed and some results are depicted in this appendix. The channel impulse response used was

$$h(n) = 0.7107\delta(n) + 0.1421\delta(n-1) + 0.2132\delta(n-2) + 0.1421\delta(n-3) + 0.6396\delta(n-4).$$

We present in Figure 17 the simulation result and upper bound for different numbers of states. As expected with 12- and 14-state trellises, the probability of error is better than in the 8-state case.

III. CONCLUSIONS AND RECOMMENDATIONS

During this period of research we made significant progress in applying the bootstrap algorithm, particularly that based on decorrelation, to two important area: separation of multi-user CDMA signals and equalization of dispersive communication channels.

We laid the ground work for the adaptive separator of a synchronous channel and examined the effect of predecisions, hard or soft limiters, on the performance. We also examined the use of the decorrelator for this separator. For the asynchronous case we presented the adaptive separator based on the one-shot approach, which turns k asynchronous users with known relative delay τ_k into $2K-1$ synchronous users. If the relative delays are unknown then

this approach can not be used. Instead a special arrangement of the decorrelator algorithm is possible. Some preliminary results were obtained using this approach. A convergence and stability analysis was also performed, and conditions were established for the synchronous separator convergence case.

Symbol-by-symbol blind channel equalization was also considered. Using the decorrelation algorithm, inter-symbol interference was reduced. Particularly, for the first time, for the decision feedback structure. The approach proves to achieve global convergence in contrast to other methods. Some results on anchoring the constant modulus were also obtained.

For blind maximum likelihood sequence estimation we introduced the idea of quantizing the channel response to be able to perform an estimation of the data sequence when the channel is unknown. A method of using two simultaneous trellises, one for data and one for channel response, was introduced and shown to result in a considerable reduction of computational complexity and it prevails with regards to the complexity/performance tradeoff.

We then introduced a method of reduced complexity sequence estimation which can be used for both a long channel response and a large constellation. It also provides more freedom in number of reduced states and gives more flexibility in design.

In summary, during this period of research, aside from adding information on issues of implementation, we have a vast amount of results in the two main topics of applications. A list of pubications resulting from this research is given in Section V.

The work carried out led us to make a number of recommendations for further study.

1. **Adaptive threshold level for the separators of multi-user CDMA signals with soft decisions**:

   It was claimed in section 2.2.2 that soft decision improves performance of the synchronous separator, particularly at a region with a low interference level. The threshold level of the soft decision was determined heuristically from the observed values of the

decorrelator outputs. We suggest to search for optimum threshold (adaptive) to obtain the best performance possible with predecisions, and hence obtain the best separator of this kind.

2. **Separator of asynchronous multi-user CDMA signals with uncertainty in relative delays:**

   Some preliminary results were obtained in section 2.2.4. However, even with the one-shot approach, the relative delays τ_k obtained from some measurement may contain some uncertainty. In this regard:

   - One may want to examine the effects of uncertainty on the performance of the separator,
   - Suggest an adaptive structure to reduce the effect of this uncertainty,
   - Examine the possibility of using this adaptive structure for handling the asynchronous multi-user case with unknown relative delays, and avoid the need for separate estimators for finding these delays.

   Preliminary estimation of these problems suggest that the bootstrapped algorithm can be advantageous in solving them.

3. **Effect of channel fading and multi-path on the separator's performance:**

   All work previously done assumed that the channel was unfaded and slowly varying. In reality, channels are stochastic and suffer from fading. Models for these channels are needed to examine their effects on performance. This is an important topic for extending current results.

4. **Multi-Carrier, multi-user, CDMA separator:**

   Especially for a multi-path channel, a complex RAKE receiver was proposed. Multi-carrier CDMA turns the wideband CDMA full channel into many narrow band chan-

nels. Therefore, multi-path becomes a flat fade of the different narrow-band channels, and hence it illuminates the need for a RAKE receiver. Again some preliminary results were obtained with this structure. The bootstrap decorrelating algorithm was proven to be crucial in its implementation.

5. **Further work on equalization using the decorrelation algorithm:**

   For most practical communication channels, the response is an FIR minimum or no minimum phase with a post- and pre-cursor. In section 2.3.1 we modeled the channel as FIR with only post-cursors and showed that the blind decision feedback equalizer can be used successfully with the decorrelation algorithm for the linear equalizer, with or without a pre-cursor remained open. More work in this direction is needed.

6. **SPW Modules:**

   Further implemention the results of this report with modules of the signal processing workstation (SPW) are needed.

IV. REFERENCES

[1] Y. Bar-Ness, A. Dinc and H. Messer-Yaron, *The bootstrapped Algorithm: Fast Algorithm for Blind Signal Separation,* Final Technical Report, RL-TR-93-24, April, 1993.

[2] Y. Bar-Ness, A. Dinc, H. Messer-Yaron, Z. Siveski, R. Kamel and Ruth Onn, *The bootstrapped Algorithm: A Fast Algorithm for Blind Signal Separation,* Final Technical Report, RL-TR-93-236, Dec., 1993.

[3] Y. Bar-Ness and J. Rokah, "Cross-Coupled Bootstrapped Interference Canceler," *Antenna and Propagation Symposium,* conference proceedings, pp. 292-295, 1981.

[4] Y. Bar-Ness, J. Carlin and M. Steinberger, "Bootstrapping Adaptive Cross Pol Cancelers for Satellite Communication," *Proc. of IEEE International Conference on Communications (ICC),*, No. 4F5, Philadelphia, PA, June 1982.

[5] W. Carlin, Y. Bar-Ness, S. Gross, M. Steinberger and W. Studdiford, "An IF Cross-Pol Canceler for Microwave radio Systems," *Journal on Selected Area in Communication - Advances in Digital Communication by Radio,* Vol. SS AC-5, No. 3, pp. 502 - 514, April, 1987.

[6] Y. Bar-Ness and J. Carlin, "Cross-Pol Canceler Architecture for Microwave Radio Applications," International Conference on Communications, *Conference Proceedings* paper No. 52.5, Seattle WA, June 1987.

[7] Y. Bar-Ness, "Effect of Number on Taps on cross-Pol Canceler Performance for Digital Radio," *The IEEE Global Communication Conference (Globecom)* 1987, paper No. 31.7.

V. LIST OF PUBLICATIONS

5.1 Papers in Professional Journals

1. Kamel, R. E. and Bar-Ness, Y., "Reduced State Sequence estimation of Digital Sequences in Dispersive Channels using State Partitioning," Electronic Letters, Vol. 30, pp. 14-17, 1994.

2. R.E. Kamel and Y. Bar-Ness, "Blind Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEE Electronics Letters* Vol. 30 No. 7, pp. 537-539, 31 March 94.

3. Siveski, Z., Bar-Ness, Y., and Chen, D. W., "Error Performance of Synchronous Multiuser Code Division Multiple Access Detector with Multidimensional Adaptive Canceler," accepted by European Transactions on Telecommunications and Related Technologies.

5.2 Papers in Conference Proceedings

1. Siveski, Z. and Bar-Ness, Y., "Two Stage Detection Scheme in Syncronous Two-User CDMA Systems," MILCOM'93.

2. Kamel, R.E. and Bar-Ness, Y., "Blind Decision Feedback Equalization Using Decorrelation Criterion," 1993 GLOBECOM Communication Theory Mini-Conference.

3. Siveski, Z., Bar-Ness, Y., and Chen, D.W., "Adaptive Signal Separation of Synchronous Code Division Multiple Access Applications," the 1994 International Zurich Seminar on Digital Communications, LNCS 783, pp. 67-75, March 1994.

4. Chen, D.W., Siveski, Z., and Bar-Ness, Y., "Synchronous Multiuser CDMA Detector with Soft Decision Adaptive Canceler." the 28th Annual Conference on Information Sciences and Systems, Princeton, NJ, March 1994.

5. Kamel, R.E. and Bar-Ness, Y., "Anchored Constant Modulus Algorithm (ACMA) for Blind Equalization," the International Conference on Communications, New Orleans, LA, May, 1994.

6. Bar-Ness, Y., Siveski, Z., and Chen, D.W., "Bootstrapped Decorrelating Algorithm for Adaptive Interference Cancelation in Synchronous CDMA Communication Systems," the IEEE Third International Symposium on Spread Spectrum Techniques and Applictions, Oulu, Finland, pp. 162–166, July 4-6, 1994.

7. Kamel, R.E. and Bar-Ness, Y., "Reduced-State Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference Using State Partitioning," the 1994 IEEE International Symposium on Information Theory, Norway, June 1994.

8. Haimovich, A., Onn, R., and Bar-Ness, Y., "Adaptive Detectors for Multi-Channel Signals in Code Division Multiple Access Systems," EUSIPCO'94, Edinburgh, Scotland, pp. 256–259, September 13-16, 1994.

9. Siveski, Z., Zhong, L., and Bar-Ness, Y., "Adaptive Multiuser CDMA Detector for Asynchronous AWGN Channels," PIMRC'94, The Hague, Netherlands, pp. 416–419, September 21-23, 1994.

10. Bar-Ness, Y., Chen, D.W., and Siveski, Z., "Adaptive Multiuser Bootstrapped Decorrelating CDMA Detector in Asynchronous Unknown Channels," PIMRC'94, The Hague, Netherlands, pp. 533–537, September 21-23, 1994.

11. Kamel, R. and Bar-Ness, Y., "Error Performance of Blind Decision Feedback Equalizer Using Decorrelation," MILCOM'94, Fort Monmouth, NJ, pp. 618–622, October 2–5, 1994.

12. Zhu, B., Ansari, N., Siveski, Z., Bar-Ness, Y., "Convergence and Stability Analysis of Synchronous Adaptive CDMA-Based PCS Receiver," MILCOM'94, Fort Monmouth, NJ, pp. 923–927, October 2-5, 1994.

13. Bateman, A., Bar-Ness, Y., and Kamel, R., "Decorrelation Algorithm for Blind Decision Feedback Equalizer with Lattice Structures," the Sixth Digital Signal Processing Workshop, Yosemite, CA, pp. 253–256, October 2-5, 1994.

5.3 Papers Submitted to Professional Journals

1. R.E. Kamel and Y. Bar-Ness, "Reduced-Complexity Sequence Estimation Using State Partitioning," submitted to *IEEE Trans. on Communications.*

2. R.E. Kamel and Y. Bar-Ness, "Blind Decision Feedback Equalization Based on Decorrelation," submitted to *European Transactions on Telecommunications.*

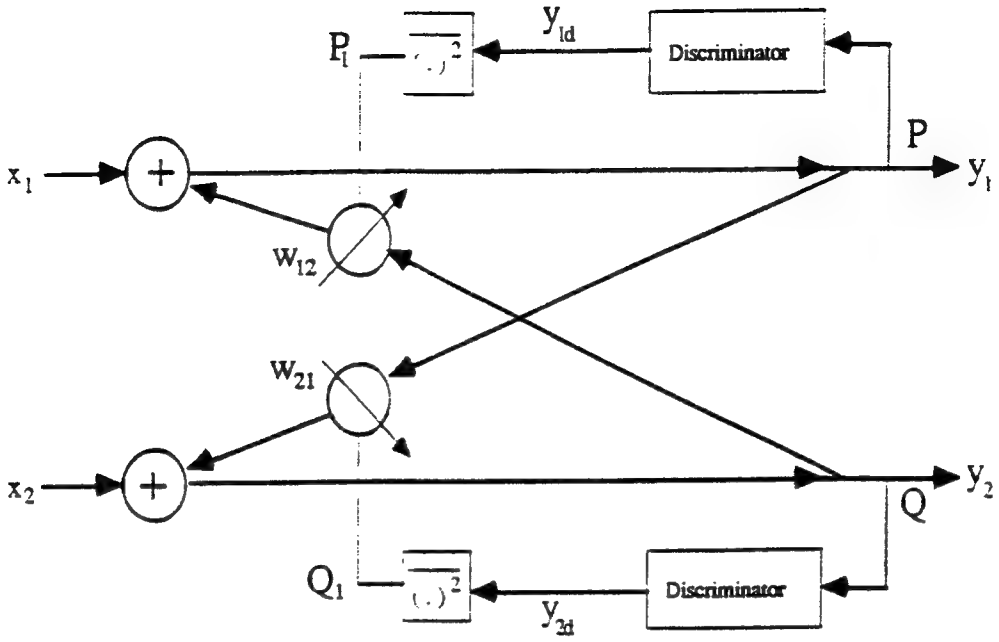


Figure 1 The backward/backward dual polarization separator.

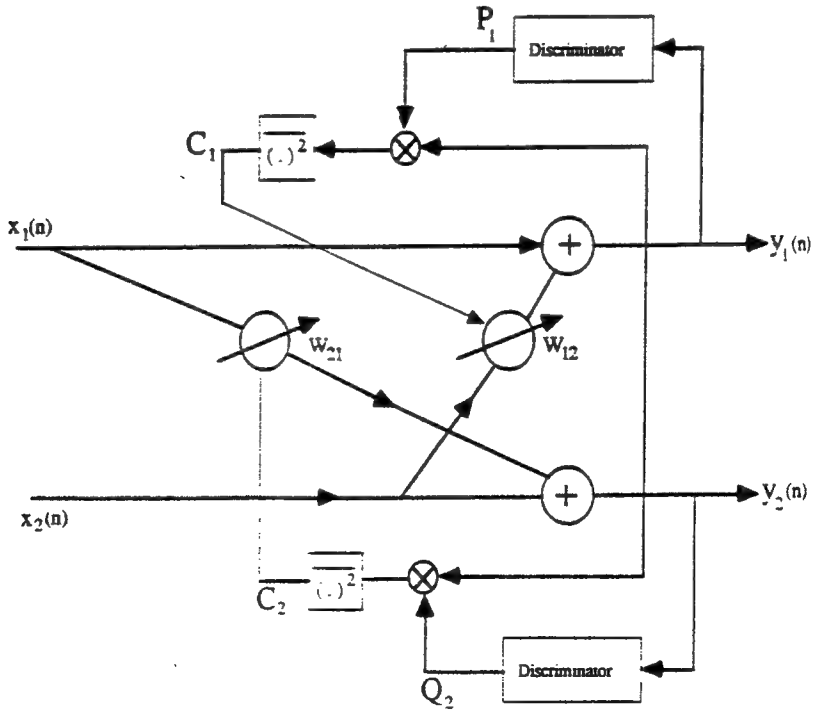


Figure 2 The forward/forward dual polarization separator.

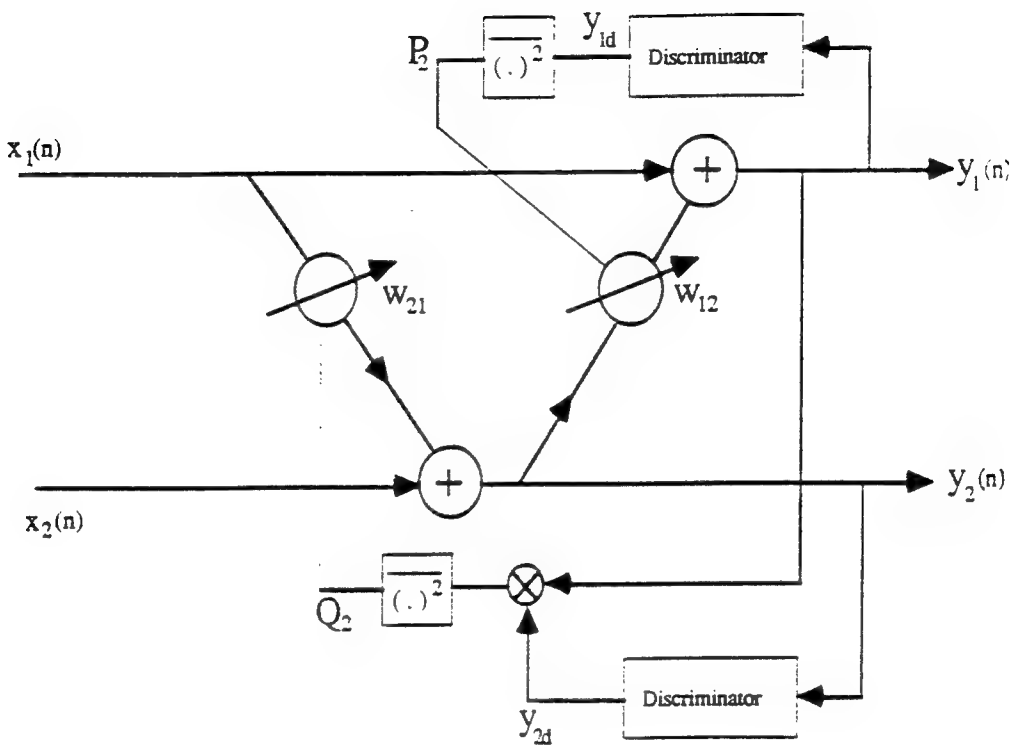


Figure 3 The forward/backward dual polarization separator.

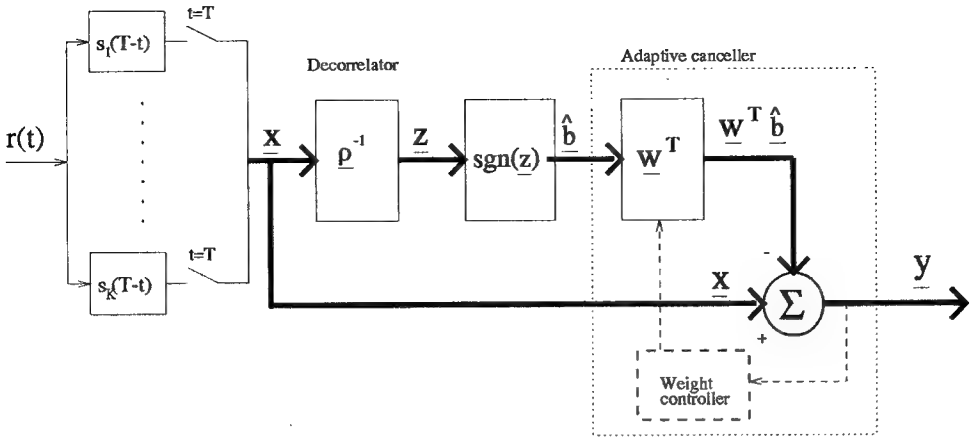


Figure 4 Adaptive Separator of Synchronous Multi-User CDMA Signals

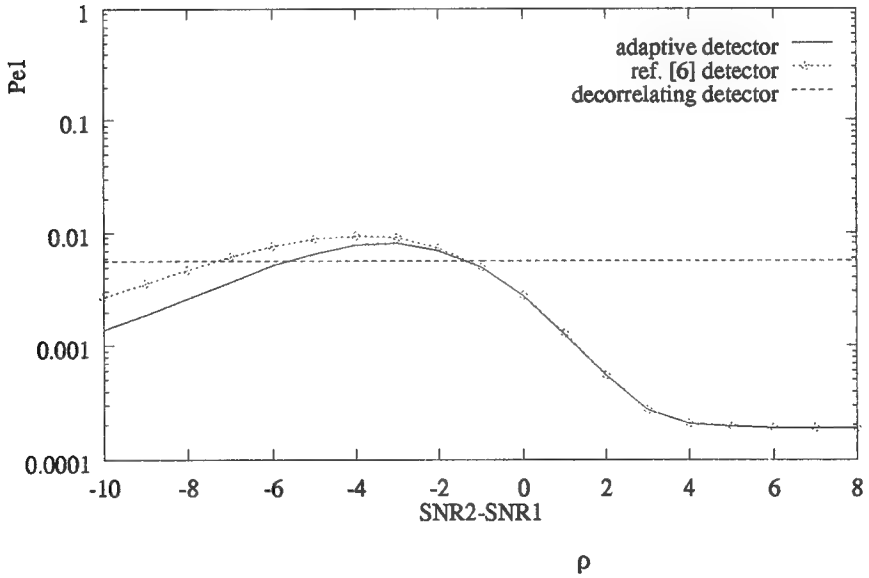


Figure 5 Error Probability of User 1 for $K = 2$, and $SNR_1 = 8$ dB and $\rho_{12} = 0.7$.

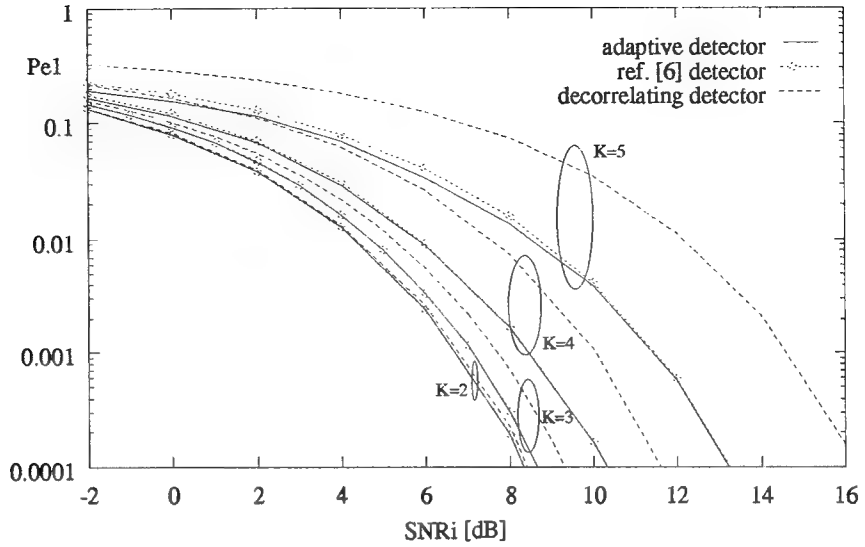


Figure 6 Error probability of User 1 for k=2 to k=5

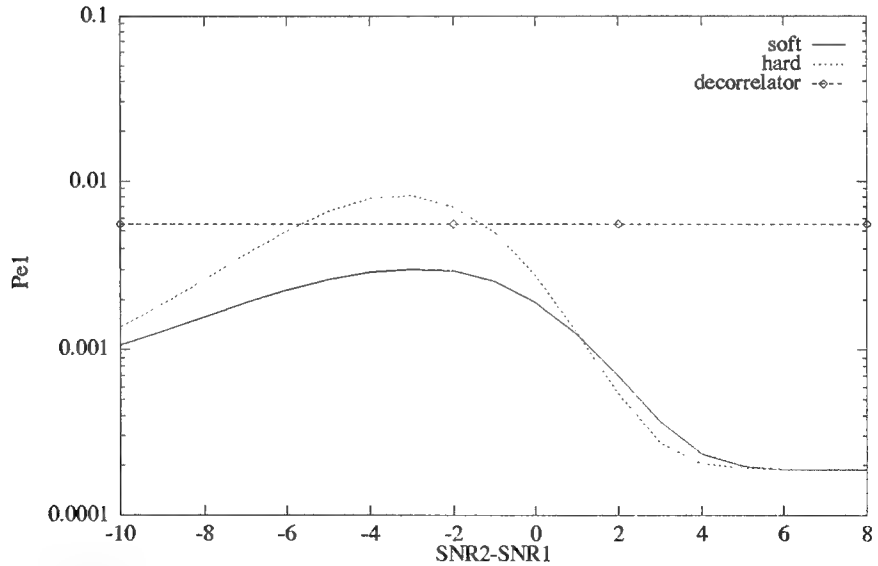


Figure 7 Effect of Soft Limiter on Adaptive Separator.
Error probability of User 1 for $K = 2$, and $SNR_1 = 8$ dB and $\rho_{12} = 0.7$.

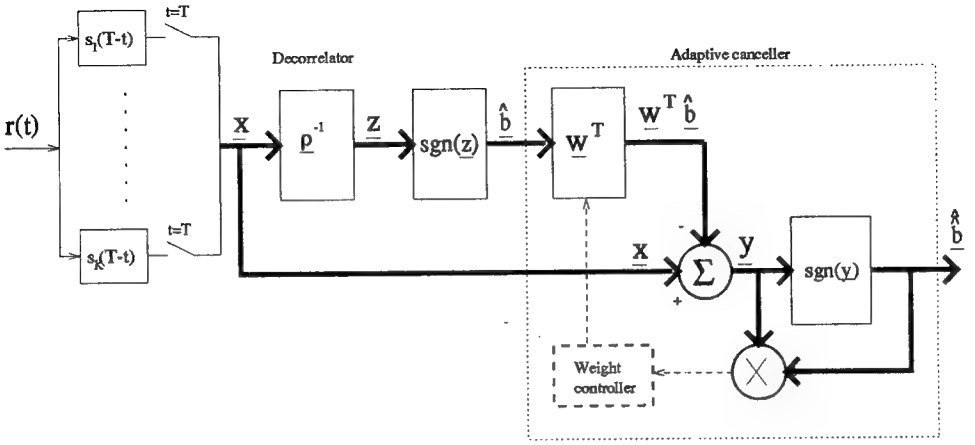


Figure 8 Adaptive Separator of Synchronous Multiuser CDMA Signal Using the Bootstrapped Decorrelating Algorithm

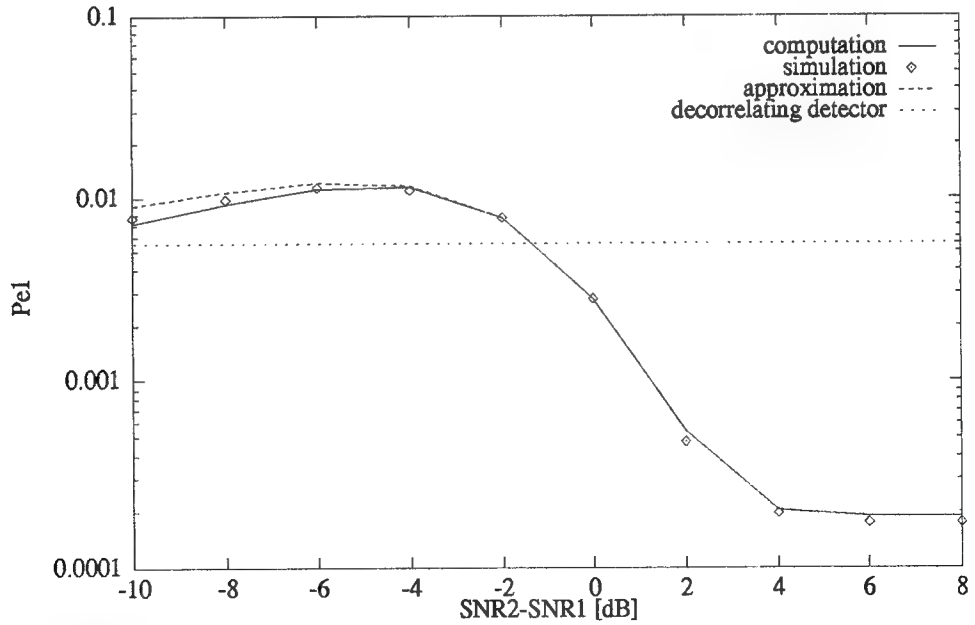


Figure 9 Performance of Separator using Decorrelating Algorithm. Probability of Error for User 1 (K=2, $SNR_1 = 8$ dB, $\rho_{12} = 0.7$)

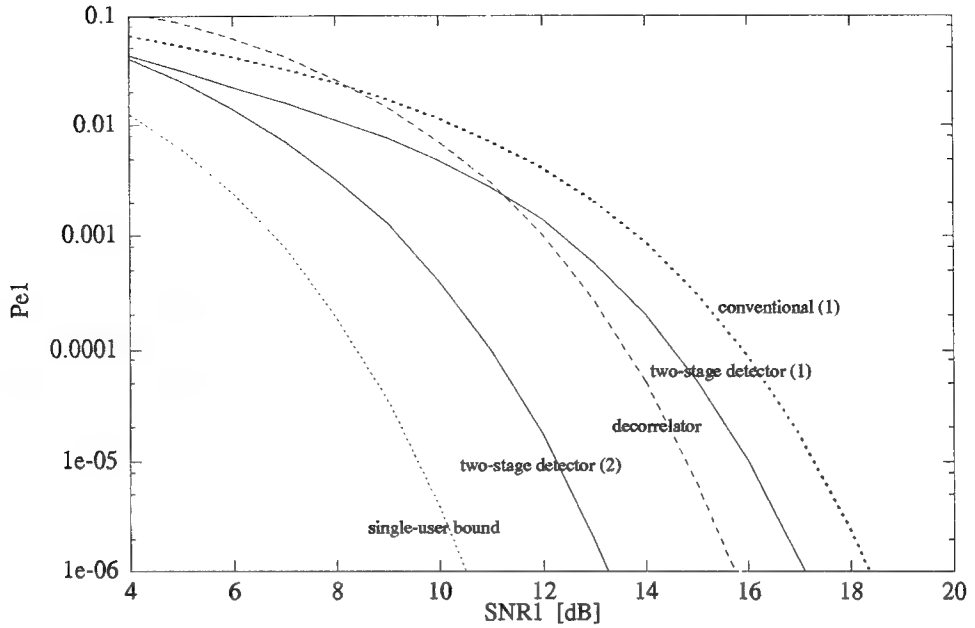


Figure 10 Performance of One-Shot Separator of Asynchronous Signals.
Error Probability of user 1 for K=2, $\rho_{12} = 0.2$, $\rho_{21} = 0.6$, $e_1 = 0.4$
(1) $a_2/a_1 = 0.6$ (2) $a_1/a_2 = 0.6$

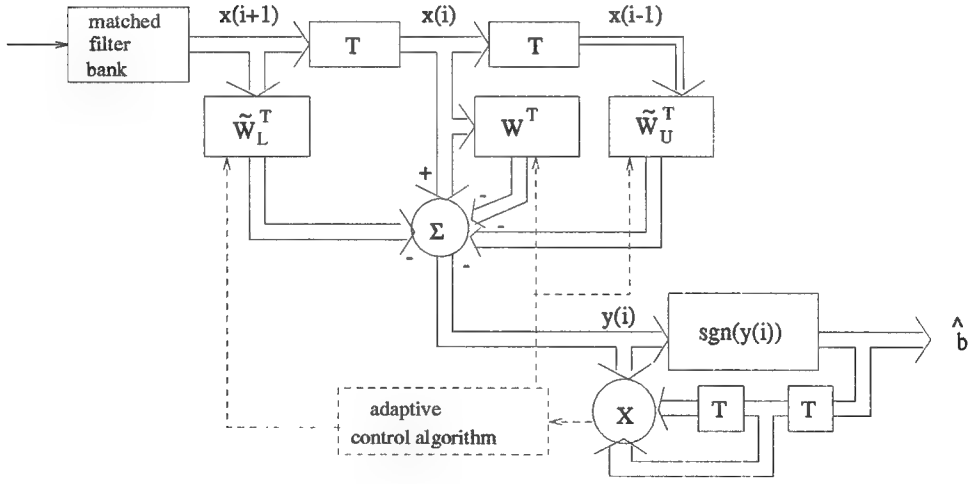


Figure 11 Proposed Bootstrapped Decorrelator for Asynchronous Unknown Channel

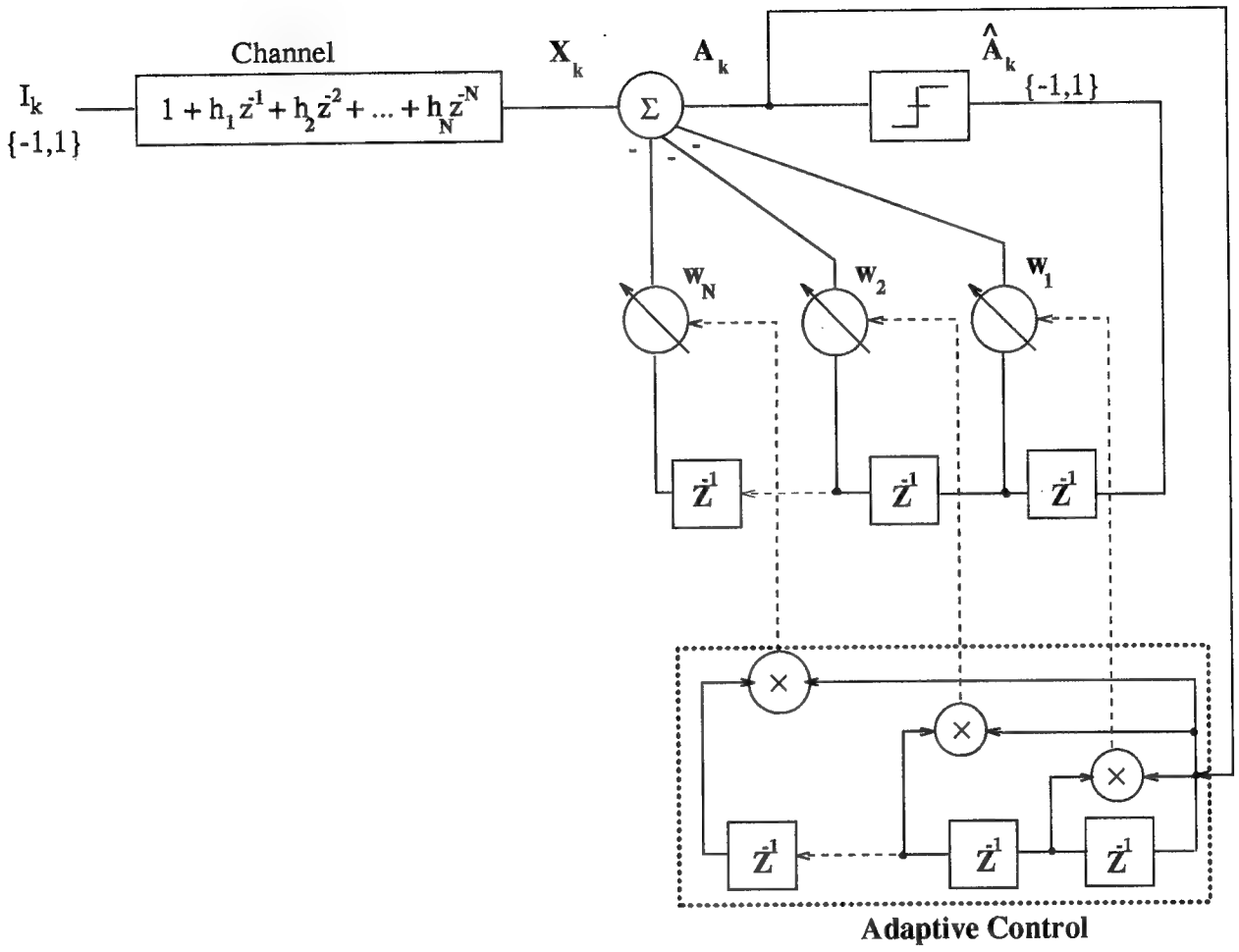


Figure 12 Decision Feedback Equalizer with Decorrelation Control

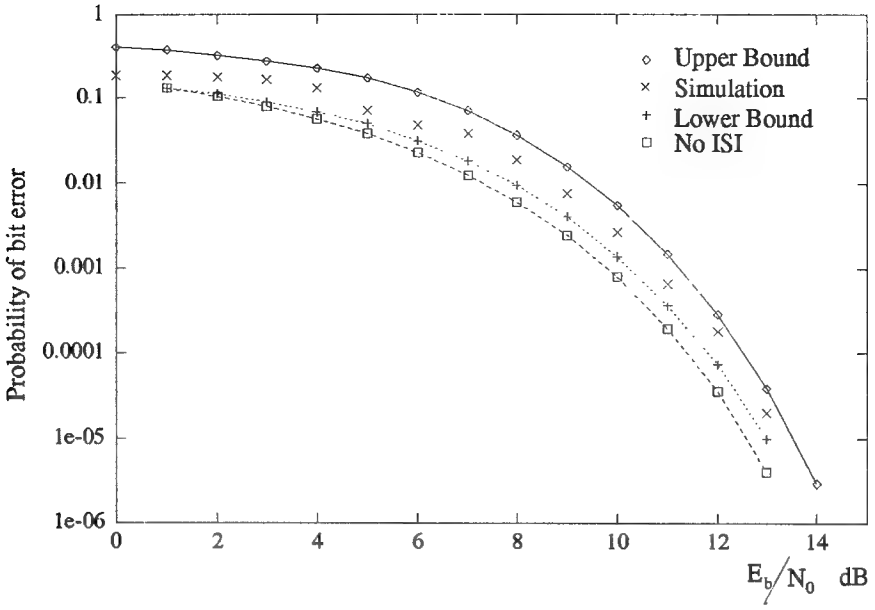


Figure 13 Performance of Decision Feedback Equalizer with Decorrelation Control.

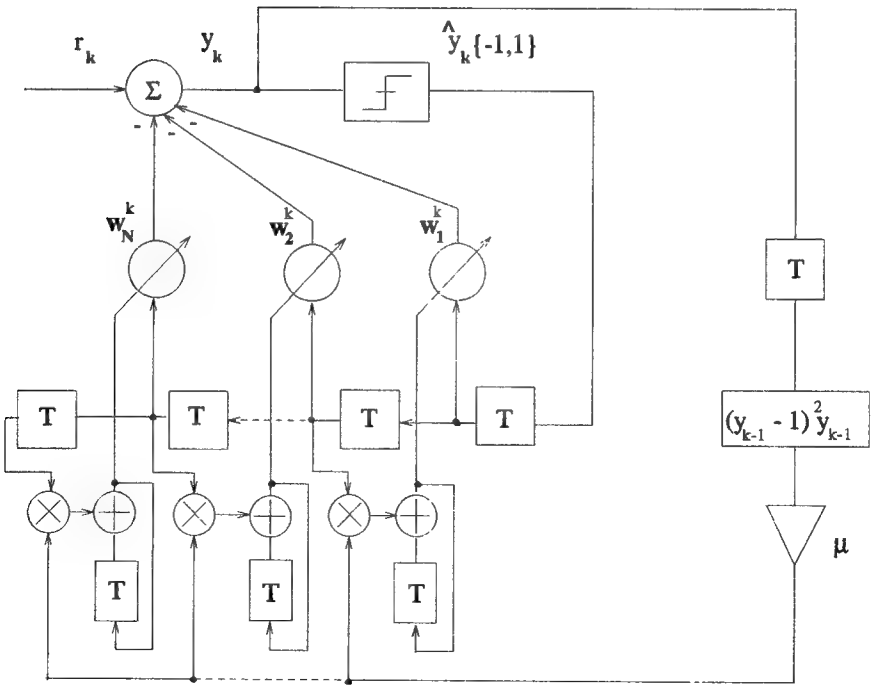


Figure 14 The Anchored DFE using the CMA

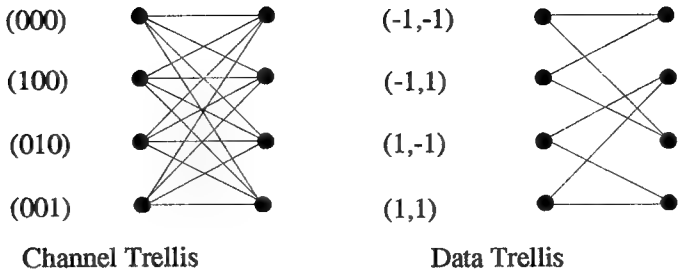


Figure 15 Channel and data trellises

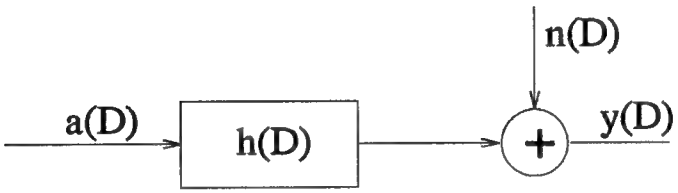


Figure 16 Discrete Channel model

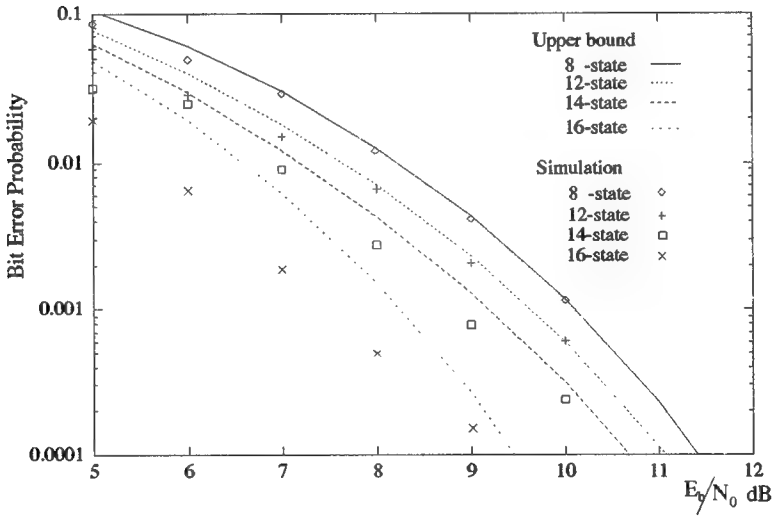


Figure 17 Performance of Reduced Complexity Sequence Estimate

APPENDIX A: PART I

FAST DECORRELATION ALGORITHMS FOR SIGNAL SEPARATION

by
Alex Haimovich, R. Manzo and Yeheskel Bar-Ness

ABSTRACT

LMS and RLS type algorithms are suggested for decorrelation of multi-channel systems outputs. These algorithms act as signal separators when applied to unknown linear combinations of the inputs. The performance of the suggested algorithms is compared with that of the conventional LMS and RLS algorithms that minimize the mean square error. It is shown that the correlation matrix eigenvalue spread associated with the LMS decorrelator is always smaller than the eigenvalue spread corresponding to the conventional LMS, resulting in faster convergence speed for the decorrelator. A new RLS type decorrelator algorithm is suggested. The RLS decorrelator is shown to be faster than the LMS decorrelator, not affected by the eigenvalue spread, and comparable in speed with the conventional RLS algorithm. Convergence analysis by simulation shows that the RLS algorithms and the LMS decorrelator have wider regions of convergence than the conventional LMS.

I. INTRODUCTION

The separation problem addressed in this paper is the recovery of unknown independent sources from observations of a linear mixture of the sources. We are concerned with a multi-channel input multi-channel output system where each output is an unknown linear combination of the inputs. This type of problem arises in numerous applications such as canceling cross-polarization interference in dual polarized systems, detection in multiple access communications systems, and separating multiple speech signals, to name a few.

One approach to signal separation is to view it as an interference cancellation problem. The signal-to-noise-and-interference ratio at each output channel is enhanced by suppressing the co-channel interference. When a reference signal is available, the mean square error (MSE) between the output and the reference signal is minimized by the classical Wiener filter, which can be implemented using steepest descent algorithms such as the LMS and RLS. For example, the LMS algorithm, operating as an interference canceler, has been suggested as a separator of cross-polarized signals [1].

Begining with the early eighties [2] a class of signal separators have been proposed that in effect estimates the parameters of the mixture between the sources. A detailed steady state analysis of three cross-coupled "noise cancelers" structure can be found in [3]. This research was aimed mainly for communication application such as satellite and digital microwave dually polarized links. It assumed that the signals to be separated are uncorrelated and showed separation, provided that "some knowledge" is available by which the signal can

discriminated. Later the same separator structures were suggested for application to neural networks [4]. Their algorithm for separation, however, assumed the signals to be statistically independent and therefore the recovered signals should be statistically independent. While it is difficult to devise criteria for testing statistical independence, it is possible to design cost functions for decorrelation between outputs. Independence cannot be achieved directly, but it can be approached by minimizing higher order cross-moments by way of gradient stochastic algorithms with costs defined by non-linear functions of the outputs [5]. These signal separators are derived from criteria different than those of the interference cancelers, furthermore, they do not require a reference signal, hence belong to the class of *blind separators*. Their operation suggests a bootstrap process where cross-correlations of the outputs, or of functions of the outputs, are used to control feedforward, or feedback, filter weights. Simulations have shown that LMS decorrelating algorithms are faster than LMS algorithms minimizing the mean square error [6].

In this paper we compare various signal separation algorithms based on the MSE and decorrelation criteria. The model and examples reflect a multi-user communication system such as CDMA. Each user transmits an assigned unique waveform modulated by BPSK data. The waveforms are assumed orthogonal and the mixture of the signals arises due to the channel effects such as crosstalk and dispersion. The main contributions of the paper are: (1) prove that the eigenvalue spread associated with the LMS decorrelating algorithm is smaller than the spread associated with the LMS, thus providing an explanation for results previously reported regarding the speed of convergence of the LMS decorrelating algorithm, (2) introduce a new decorrelating RLS type algorithm.

The system model is presented in Section . In Section we define separation criteria and based on them compute expressions for the steady state weight vectors. Adaptive algorithms and the con
rgence analysis of the LMS decorrelator are developed in Section . In Section the various algorithms are compared by computer simulations. Conclusions are provided in Section .

II. PROBLEM STATEMENT

Consider the communication of N sources through N channels. We make the following basic assumptions:

1. The sources are independent.
2. The outputs of the channels consist of a linear mixture of the inputs.
3. Number of inputs is equal to the number of sources.

The observation signals $x_n(t)$ are linear combinations of the input sources, such that:

$$x_m(t) = \sum_i \sum_{n=1}^{N} a_{mn} b_n^{(i)} \sqrt{\xi_n} h_n(t) + \nu(t) \tag{1}$$

for $1 \leq m \leq N$. $b_n^{(i)} \in \{-1, 1\}$ is the n-th information symbol bit at the i-th time interval. The information symbols are assumed to be statistically independent and equiprobable.

$h_n(t)$ is the received pulse signature associated with the n-th user. The energy of $h_n(t)$ is normalized to 1, i.e. $\int_o^{T_s} h_n(t)\,\mathrm{d}\,t = 1$, and ξ_n is the energy of the received signal from the n-th user. Cross-correlations between the sequences are included in the matrix $\mathbf{A}$. $\nu(t)$ is (possibly correlated) additive Gaussian noise. The coupling fixed parameters a_{mn} are not known. The received signals after match filtering and sampling may be represented by the vectors:

$$\mathbf{x} = \mathbf{AEb} + \mathbf{v} \tag{2}$$

where $\mathbf{A}$ is the mixture matrix, $\mathbf{E}$ is a diagonal matrix, $\operatorname{diag}\mathbf{E} = \left[\sqrt{\xi_1}, \cdots, \sqrt{\xi_N}\right]^T$, and $\mathbf{b} = [b_1, \cdots, b_N]^T$ is the vector of information bits. The time superscript has been dropped since the decision on $b^{(i)}$ requires observation of signals during that interval only, thus $\mathbf{b}$ represents the vector of information bits at any time. The vector $\mathbf{v}$ consists of independent sample
mean Gaussian noise with covariance matrix $\mathrm{E}\left[\mathbf{v}\mathbf{v}^T\right] = \sigma_v^2\mathbf{G}$, where $\mathbf{G}$ is the known matrix of the noise cross-correlations between the outputs of the match filters.

III. SIGNAL SEPARATION CRITERIA

Two signal separation criteria are considered: the Mean Square Error (MSE) separation and the signal decorrelation criterion. The criteria are used to develop control algorithms for the network weights. We will consider separation of the n-th user signal from the rest of the signals, other channels being treated similarly. To yield the n-th user ou
put y_n, the input vector $\mathbf{x}$ is applied to the adaptive filter represented by the $N \times 1$ column vector $\mathbf{w}_n$. For BPSK transmission a sign test can be used for detecting the data bit: $b_n = \operatorname{sgn} y_n$.

3.1 MSE Separation

The MSE signal separator minimizes the mean squared error between its output and a reference signal. Typically the reference is initially supplied by a training signal. When the adaptive weights converged and the errors with respect to the training signal are small, the detector is switched to operate in decision directed mode, with the reference signal being supplied by the estimated symbol. The MSE signal separator is in effect the optimum linear detector for the n-th user. It acts as a canceler for the co-channel interference from the other users and minimizes the error between the reference signal and the output y_n. In decision directed mode the reference signal is supplied by the estimated data symbol, hence we have $e(k) = \widehat{b}_n(k) - y_n(k)$. The MSE is defined $\epsilon = \mathrm{E}\left[e^2(k)\right]$. In the following we drop the explicit time dependency since signals are assumed stationary. The steady state weight vector for separating the n-th user signal is given by the well known Wiener-Hopf equation,

$$\mathbf{w}_n = \mathbf{R}_x^{-1}\mathbf{r}_{\widehat{b}_n x} \tag{3}$$

where $\mathbf{R}_x = E\left[\mathbf{x}\mathbf{x}^T\right]$ is the input correlation matrix, and $\mathbf{r}_{\widehat{b}_n x} = E\left[\widehat{b}_n\mathbf{x}\right]$ is the cross-correlation between the input and the estimated symbol. We will show that in the absence

of noise the MSE criterion indeed leads to signal separation. When the input vector $\mathbf{x}$ is given by the model in equation 2, the input correlation matrix can be written

$$\begin{aligned} \mathbf{R}_x &= \mathbf{A}\mathbf{E}\,\mathrm{E}\left[bb^T\right]\mathbf{E}\mathbf{A}^T \\ &= \mathbf{A}\mathbf{E}^2\mathbf{A}^T \end{aligned} \tag{4}$$

The cross-correlation vector is

$$\begin{aligned} \mathbf{r}_{\widehat{b}_n x} &= \mathrm{E}\left[\widehat{b}_n\left(\mathbf{A}\mathbf{E}\mathbf{b}+\nu\right)\right] \\ &= \mathbf{A}\mathbf{E}\mathbf{u}_n \end{aligned} \tag{5}$$

where $\mathbf{u}_n$ is a unit vector with the n-th element equal to 1. To make the analysis mathematically tractable it is assumed that in the decision directed mode correct decisions are made with high probability such that $\mathrm{E}\left[\widehat{b}_n\mathbf{b}\right] = \mathbf{u}_n$, and in general $\mathrm{E}\left[\widehat{\mathbf{b}}\mathbf{b}\right] = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. From , (4) and (5) the weight vector $\mathbf{w}_n$ is equal,

$$\mathbf{w}_n = \mathbf{A}^{-T}\mathbf{E}^{-1}\mathbf{u}_n \tag{6}$$

.

The information bit is recovered when this weight vector is applied to the noiseless data: $y_n = \mathbf{w}_n^T\mathbf{x} = \mathbf{u}_n^T\mathbf{E}^{-1}\mathbf{A}^{-1}\mathbf{A}\mathbf{E}\mathbf{b} = b_n$.

3.2 Decorrelation Signal Separation

The decorrelation separator is different from the MSE separator. The MSE weight vector for any user is controlled by the error derived solely from the output corresponding to that user. In contrast, the decorrelator weight vector is controlled by *all* the channels' outputs. Suggesting an operation where each channel is helped by the other channels this structure has been referred to as *bootstrapped* [6]. The decorrelator seeks to decorrelate the output y_n from a set of reference signals which represents the other users' symbols. Operating in a mode similar to the decision directed mode, decorrelation is achieved when

$$\mathrm{E}\left[y_n\widehat{\mathbf{b}}\right] = c_n\mathbf{u}_n \tag{7}$$

where c_n is a constant. When condition (7) is met, y_n is decorrelated from all components of $\widehat{\mathbf{b}}$ other than its own b_n. First we will show that in the noiseless case this criterion indeed leads to signal separation, then we will define and optimize a cost function to compute the steady state decorrelating weight vector.

For the noiseless case we have $\mathrm{E}\left[y_n\widehat{\mathbf{b}}\right] = \mathrm{E}\left[\left(\mathbf{w}_n^T\mathbf{A}\mathbf{E}\mathbf{b}\right)\widehat{\mathbf{b}}\right]$. After a few algebraic manipulations, we get that (7) is equivalent to $\mathbf{w}_n^T\mathbf{A}\mathbf{E}\mathbf{u}_n = c_n\mathbf{u}_n$, for $1 \le n \le N$. If we define the matrices $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_N]$ and $\mathbf{C} = \mathrm{diag}\,[c_1, \cdots, c_N]$, then the weight matrix $\mathbf{W}$ can be written:

$$\mathbf{W} = \mathbf{C}\left(\mathbf{A}^T\mathbf{E}\right)^{-1} \tag{8}$$

Indeed, when this weight matrix is applied to the input $\mathbf{x}$, the output is proportional to the data vector: $\mathbf{y} = \mathbf{W}^T\mathbf{x} = \mathbf{W}^T\mathbf{A}\mathbf{E}\mathbf{b} = \mathbf{C}\mathbf{b}$.

In the general case when noise is present, the decorrelation condition in (7) can be achieved by defining the cost function $J = \mathbf{w}_n^T \mathbf{R}_{\widehat{bx}} \mathbf{w}_n$, where $\mathbf{R}_{\widehat{bx}} = \mathrm{E}\left[\widehat{\mathbf{b}}\mathbf{x}^T\right]$, and minimizing it with respect to $\mathbf{w}_n^T$ and subject to the condition $\mathbf{w}_n^T \mathbf{u}_n = 1$. The condition sets $w_{nn} = 1$, such that the n-th component of the input is transferred directly to the output. Indeed the minimum of the cost function J is achieved when the gradient of J is set to zero and $w_{nn} = 1$:

$$\nabla J = \mathbf{R}_{\widehat{bx}} \mathbf{w}_n = \mathrm{E}\left[y_n \widehat{\mathbf{b}}\right] = c_n \mathbf{u}_n \tag{9}$$

which is equivalent to the decorrelation condition (7). The steady state decorrelating weight vector is given by:

$$\mathbf{w}_n = \beta \mathbf{R}_{\widehat{bx}}^{-1} \mathbf{u}_n \tag{10}$$

where the scaling factor $\beta = \left(\mathbf{u}_n^T \mathbf{R}_{\widehat{bx}}^{-1} \mathbf{u}_n\right)^{-1}$ is set to meet the linear constraint.

3.3 Adaptive Algorithms

When the signal environment is not stationary weight vectors for the MSE and the decorrelator can be calculated using steepest descent algorithms.

3.3.1 MSE Separator Algorithms

The LMS algorithm for updating the weight vector of the MSE separator is given by

$$\begin{aligned} \mathbf{w}_n(k) &= \mathbf{w}_n(k-1) + \mu e(k) \mathbf{x}(k) \\ e(k) &= \widehat{b}_n(k) - \mathbf{x}^T(k) \mathbf{w}_n(k-1) \end{aligned} \tag{11}$$

To distinguish it from the LMS decorrelator we refer to (11) as the LMS error algorithm. The convergence analysis of the LMS algorithm is well known. The necessary and sufficient condition for the convergence of the LMS algorithm is that the convergence factor $0 < \mu < 2/\lambda_{\max}(\mathbf{R}_x)$, where $\lambda_{\max}(\mathbf{R}_x)$ is the largest eigenvalue of the covariance matrix $\mathbf{R}_x$. It is also well known that the speed of the LMS algorithm is determined by eigenvalue spread $\chi = \lambda_{\max}(\mathbf{R}_x) / \lambda_{\min}(\mathbf{R}_x)$.

The RLS algorithm for the MSE separator is based on a recursive implementation of relation (3) using the following updates of the covariance matrix $\mathbf{R}_x$ and cross-correlation vector $\mathbf{r}_{\widehat{b}_n x}$:

$$\begin{aligned} \mathbf{R}_x(k) &= (1-\alpha)\mathbf{R}_x(k-1) + \alpha \mathbf{x}(k)\mathbf{x}^T(k) \\ \mathbf{r}_{\widehat{b}_n x}(k) &= (1-\alpha)\mathbf{r}_{\widehat{b}_n x}(k-1) + \alpha \mathbf{x}(k)\widehat{b}_n(k) \end{aligned} \tag{12}$$

where α is the forgetting factor. From (12) and (3) we get the following recursion relations:

$$\begin{array}{rcl} \mathbf{g}(k) & = & \frac{\mathbf{R}_x^{-1}(k-1)\mathbf{x}(k)}{\frac{1-\alpha}{\alpha}+\mathbf{x}^T(k)\mathbf{R}_x^{-1}(k-1)\mathbf{x}(k)} \\ \mathbf{R}_x^{-1}(k) & = & \frac{1}{1-\alpha}\left[\mathbf{I}-\mathbf{g}(k)\,\mathbf{x}^T(k)\right]\mathbf{R}_x^{-1}(k-1) \\ e(k) & = & \alpha\widehat{b}_n(k)-\mathbf{x}^T(k)\,\mathbf{w}_n(k-1) \\ \mathbf{w}_n(k) & = & \mathbf{w}_n(k-1)+\mathbf{g}(k)\,e(k) \end{array} \tag{13}$$

3.3.2 LMS Decorrelator

The LMS decorrelator is a steepest descent algorithm that seeks to null the instantaneous estimate of the gradient of the cost function J. From (10) we have, $\nabla J = \mathbf{R}_{\widehat{bx}}\mathbf{w}_n$, and for the instantaneous estimate $\widehat{\nabla} J(k) = \widehat{\mathbf{b}}(k)\,\mathbf{x}^T(k)\,\mathbf{w}_n(k) = y_n(k)\,\mathbf{w}_n(k)$. According to our formulation $w_{nn} = 1$, hence to preclude adaptation of w_{nn} we premultiply the gradient by $\mathbf{U}_n = \left(\mathbf{I}-\mathbf{u}_n\mathbf{u}_n^T\right)$ which zeros its n-th component. The LMS decorrelator algorithm is given by:

$$\begin{array}{rcl} \mathbf{w}_n(k) & = & \mathbf{w}_n(k-1)-\mu y_n(k)\,\mathbf{U}_n\widehat{\mathbf{b}}(k) \\ w_{nn} & = & 1 \end{array} \tag{14}$$

This relation can also be interpreted directly as seeking to minimize the instantaneous correlation between the output of the separating filter for the n-th source and the other detected sources.

3.3.3 Convergence Analysis of LMS Decorrelator

The LMS decorrelator have been observed to converge faster than the conventional LMS error algorithm [6]. In this part we will prove that the eigenvalue spread for the LMS decorrelator is smaller than for the LMS error, hence the former has faster convergence speed

First we determine necessary and sufficient conditions for convergence of the LMS decorrelator. Define $\widehat{\mathbf{B}} = \left(\mathbf{I}-\mathbf{u}_n\mathbf{u}_n^T\right)\widehat{\mathbf{b}}$, we can rewrite (14)

$$\begin{array}{rcl} \mathbf{w}_n(k) & = & \mathbf{w}_n(k-1)-\mu y_n(k)\,\widehat{\mathbf{B}}(k) \\ & = & \left[\mathbf{I}-\mu\widehat{\mathbf{B}}(k)\,\mathbf{x}^T(k)\right]\mathbf{w}_n(k-1) \end{array} \tag{15}$$

B

To analyze the convergence speed we will estimate the eigenvalue spread for the LMS decorrelator and compare it with the spread for the LMS error. Using the assumptions stated previously for the LMS error analysis, we can write:

$$\begin{array}{rcl} \mathrm{E}\left[\mathbf{w}_n(k)\right] & = & \left[\mathbf{I}-\mu\mathbf{R}_{\widehat{B}x}\right]\mathrm{E}\left[\mathbf{w}_n(k-1)\right] \\ & = & \left[\mathbf{I}-\mu\mathbf{U}_n\mathbf{E}\mathbf{A}^T\right]\mathrm{E}\left[\mathbf{w}_n(k-1)\right] \end{array} \tag{16}$$

and used the relations

$$\mathbf{R}_{\widehat{B}x} = \mathbf{U}_n\mathbf{R}_{\widehat{bx}} = \mathbf{U}_n\,\mathrm{E}\left[\widehat{\mathbf{b}}\left(\mathbf{b}^T\mathbf{E}\mathbf{A}^T+\mathbf{v}\right)\right] = \mathbf{U}_n\,\mathrm{E}\left[\widehat{\mathbf{b}}\mathbf{b}^T\right]\mathbf{E}\mathbf{A}^T = \mathbf{U}_n\mathbf{E}\mathbf{A}^T. \tag{17}$$

Without loss of generality, consider the separation of source $n = 1$, from the other sources. Define the following matrix and vector partitions: $\mathbf{w}^T = \begin{pmatrix} w_1 & \vdots & \mathbf{W}^T \end{pmatrix}$, $\mathbf{I} = \begin{pmatrix} 1 & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & \mathbf{I}' \end{pmatrix}$, $\widehat{\mathbf{B}}^T = \begin{pmatrix} 0 & \vdots & \widehat{\mathbf{B}}'^T \end{pmatrix}$, $\mathbf{x}^T = \begin{pmatrix} x_1 & \vdots & \mathbf{X}^T \end{pmatrix}$. Then we can write equation (15) as follows:

$$\mathrm{E}\left[\mathbf{W}(k)\right] = \left[\mathbf{I}' - \mu \mathbf{R}_{\widehat{B}'x}\right] \mathrm{E}\left[\mathbf{W}(k-1)\right] - \mu \, \mathrm{E}\left[\widehat{\mathbf{B}}'(k)\, x_1(k)\, w_1(k-1)\right] \tag{18}$$

We deduce that the convergence speed of the LMS decorrelator is controlled by the spread of the eigenvalues of $\mathbf{R}_{\widehat{B}'x}$.

Proposition 1: The following relation exists between the largest eigenvalues of $\mathbf{R}_{\widehat{B}'x}$ *and* $\mathbf{R}_{\widehat{bx}}$:

$$\lambda_{\max}\left(\mathbf{R}_{\widehat{bx}}\right) \geq \lambda_{\max}\left(\mathbf{R}_{\widehat{B}'x}\right) \tag{19}$$

Proof: The matrix $\mathbf{R}_{\widehat{B}'x}$ can be expressed as a partition of the matrix $\mathbf{R}_{\widehat{bx}}$:

$$\mathbf{R}_{\widehat{bx}} = \mathbf{E}\mathbf{A}^T = \begin{pmatrix} \sqrt{\xi_1} & \vdots & \mathbf{r}_1^T \\ \cdots & \vdots & \cdots \\ \mathbf{r}_1 & \vdots & \mathbf{R}_{\widehat{B}'x} \end{pmatrix} \tag{20}$$

Define the ratio

$$\rho(z) = \frac{\mathbf{z}^T \mathbf{R}_{\widehat{bx}} \mathbf{z}}{\mathbf{z}^T \mathbf{z}} \tag{21}$$

According to the Courant-Fisher theorem [7], $\lambda_{\max}\left(\mathbf{R}_{\widehat{bx}}\right) = \max_{z} \rho(z)$. Using the matrix partition in (20) and maximizing $\rho(z)$ over the restriction $z_1 = 0$, we get

$$\lambda_{\max}\left(\mathbf{R}_{\widehat{bx}}\right) \geq \max_{z,\ z_1=0} \rho(z) = \lambda_{\max}\left(\mathbf{R}_{\widehat{B}'x}\right) \tag{22}$$

Proposition 2: The following relation exists between the smallest eigenvalues of $\mathbf{R}_{\widehat{B}'x}$ *and* $\mathbf{R}_{\widehat{bx}}$:

$$\lambda_{\min}\left(\mathbf{R}_{\widehat{bx}}\right) \leq \lambda_{\min}\left(\mathbf{R}_{\widehat{B}'x}\right) \tag{23}$$

Proof: According to a corollary to the Courant-Fisher theorem, and noting that $\mathbf{R}_{\widehat{B}'x}$ is the $(N-1) \times (N-1)$ leading principal submatrix of $\mathbf{R}_{\widehat{bx}}$, we have:

$$\lambda_{\min}\left(\mathbf{R}_{\widehat{bx}}\right) \leq \lambda_{N-1}\left(\mathbf{R}_{\widehat{B}'x}\right) = \lambda_{\min}\left(\mathbf{R}_{\widehat{B}'x}\right) \tag{24}$$

Proposition 3: If the noise is negligible then the eigenvalues of $\mathbf{R}_x$ *are equal to the square of the eigenvalues of* $\mathbf{R}_{\widehat{bx}}$.

Proof: We have the relation $\mathbf{R}_{\widehat{bx}} = \mathbf{E}\mathbf{A}^T$, and for negligible noise we have $\mathbf{R}_x = \mathbf{A}\mathbf{E}\mathbf{E}\mathbf{A}^T$. Consequently $\mathbf{R}_x = \mathbf{R}_{\widehat{bx}}^T \mathbf{R}_{\widehat{bx}}$. It follows that $\lambda_i(\mathbf{R}_x) = \lambda_i^2\left(\mathbf{R}_{\widehat{bx}}\right)$ for $1 \leq i \leq N$. In particular,

$$\chi(\mathbf{R}_x) = \frac{\lambda_{\max}(\mathbf{R}_x)}{\lambda_{\min}(\mathbf{R}_x)} = \left(\frac{\lambda_{\max}\left(\mathbf{R}_{\widehat{bx}}\right)}{\lambda_{\min}\left(\mathbf{R}_{\widehat{bx}}\right)}\right)^2 = \chi^2\left(\mathbf{R}_{\widehat{bx}}\right) \tag{25}$$

Proposition 4: The eigenvalue spread $\chi(\mathbf{R}_x)$ associated with the LMS error algorithm, is larger than the spread $\chi\left(\mathbf{R}_{\widehat{B}'x}\right)$ associated with the LMS decorrelator.

Proof: From *Propositions* 1,2 and 3 we have:

$$\chi(\mathbf{R}_x) = \chi^2\left(\mathbf{R}_{\widehat{bx}}\right) \geq \chi\left(\mathbf{R}_{\widehat{B}'x}\right) \tag{26}$$

Proposition 4 provides the explanation why the LMS decorrelator algorithm is faster than the LMS error algorithm.

3.4 RLS Decorrelator

We introduce a new decorrelating RLS type algorithm. Paralleling the development of the RLS algorithm from the MSE weight vector in (3), we developed the following RLS algorithm for the decorrelator weight vector in (10):

$$\begin{aligned}
\mathbf{g}(k) &= \frac{\mathbf{R}_x^{-1}(k-1)\widehat{\mathbf{b}}(k)}{\frac{1-\alpha}{\alpha} + \mathbf{x}^T(k)\mathbf{R}_x^{-1}(k-1)\widehat{\mathbf{b}}(k)} \\
\mathbf{R}_x^{-1}(k) &= \frac{1}{1-\alpha}\left[\mathbf{I} - \mathbf{g}(k)\mathbf{x}^T(k)\right]\mathbf{R}_x^{-1}(k-1) \\
\mathbf{v}_n(k) &= \mathbf{w}_n(k-1) - \mathbf{g}(k)\mathbf{x}^T(k)\mathbf{w}_n(k-1) \\
\mathbf{w}_n(k) &= \frac{\mathbf{v}_n(k)}{\mathbf{u}_n^T\mathbf{v}_n(k)}
\end{aligned}$$

where the last relation is designed to meet the linear constraint $\mathbf{w}_n^T\mathbf{u}_n = 1$. The four algorithms are evaluated in the next section.

IV. PERFORMANCE EVALUATION

The performance of separation algorithms was evaluated in terms of the probability of error for detecting the transmitted data. Consider the probability of bit error for the n-th user. Since the data bit can take on one of two equiprobable values the total probability of error is equal to the conditional probability $P\left[\widehat{b}_n > 0 \mid b_n = -1\right]$. The probability of error at the output of a separator with weight vector $\mathbf{w}_n$ is given by:

$$\begin{aligned}
P_e &= P\left[\widehat{b}_n > 0 \mid b_n = -1\right] \\
&= P\left[y_n > 0 \mid b_n = -1\right] \\
&= P\left[\mathbf{w}_n^T\mathbf{x} > 0 \mid b_n = -1\right] \\
&= E_{\substack{b\in\{1,-1\}^N \\ b_n=-1}}\left\{P\left[\mathbf{w}_n^T\mathbf{x} > 0 \mid b_n = -1\right]\right\} \\
&= 2^{1-N}\sum_{\substack{b\in\{-1,1\}^N \\ b_n=-1}} Q\left(\frac{t_{nn} - \sum_{j\neq n} t_{nj}b_j}{\sigma_v\sqrt{\mathbf{w}_n^T\mathbf{w}_n}}\right)
\end{aligned}$$

where $Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{y^2}{2}} dy$, t_{nj} is the j-th element of the vector $\mathbf{t}_n = \mathbf{w}_n^T \mathbf{AE}$, and the zero-mean Gaussian noise was assumed white.

The convergence of the algorithms was studied by simulations. The simulations consisted of a $N = 4$ users system. Without loss of generality, the error probability was computed for the first user. In generating the signal model we assumed a fixed cross-correlation between any two users of $a_{ij} = 0.15$, for any $i \neq j$. In the following figures the SNR is defined with respect to the first user, and the signal-to-signal ratio (SSR) is defined as the ratio of the bit energy of the first user to the bit energy of any of the other users. Each curve is the result of averaging ten independent runs. Figure 1 shows the average learning curves of the probability of error for SNR = 8 dB and SSR = - 10 dB. The LMS error does not converge, and the RLS error and RLS decorrelator converge faster than the LMS decorrelator. In Figure 2, the interference power is reduced to SSR = -5 dB, the LMS error has the slowest convergence while the other three algorithms converge at comparable speeds. Comparison of Figures 1 and 2 substantiates our claims that the LMS decorrelator is less sensitive to the eigenvalue spread than the LMS error.

To estimate the convergence region of each algorithm we took the approach suggested in [8]. A figure of merit γ is defined that relates the initial and final probabilities of error, P_{ei}, and P_{ef}, as follows:

$$\gamma = 1 - \frac{P_{ef}}{P_{ei}} \tag{27}$$

when $P_{ef} \ll P_{ei}$, $\gamma \approx 1$. Note that $\gamma = 0$ corresponds to no convergence is indicated by $P_{ef} = P_{ei}$. Figure 3 shows the convergence curves for SSR = -10 dB. The SNR is varied from -10 dB to +10 dB to result in the initial probabilities of error indicated on the abscissa. The LMS error does not converge for these parameters while the LMS decorrelator and the RLS error show slightly wider regions of convergence than the RLS decorrelator. In Figure 4 the convergence is shown for SSR = -5 dB. The RLS decorrelator shows the poorest convergence. The LMS decorrelator and the RLS error have wider regions of convergence than the LMS error.

V. CONCLUSIONS

In this paper we showed that the LMS decorrelator algorithm is faster than the LMS error algorithm due to a lesser eigenvalue spread. We introduced a new RLS type algorithm for decorrelation. The RLS decorrelator was shown to be faster than the LMS type algorithms and of comparable speed with the conventional RLS error. Regions of convergence for the LMS decorrelator and the RLS algorithms were shown to be wider than for the LMS error algorithm.

References

[1] M. Kavehrad, "Performance of cross-polarized QAM signals over non-dispersive fading channels," *AT&T Bell Lab. Technical Journal*, vol. 63, pp. 499–521, mar 1984.

[2] Y. Bar-Ness and J. Rukach, "Cross-coupled bootstrapped interference canceler," *AP-S International Symposium, Conference Proceedings*, pp. 292–295, jun 1981.

[3] Y. Bar-Ness, "Bootstrapped cross-pol interference cancelling techniques-steady state analysis," *Bell Lab Technical Report*, 1982.

[4] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetric architecture," *Signal Processing, Elsevier*, vol. 24, pp. 1–10, jul. 1991.

[5] C. Jutten and J. Herault, "Blind separation of sources, Part II: Problems statement," *Signal Processing, Elsevier*, vol. 24, pp. 11–20, jul. 1991.

[6] A. Dinc and Y. Bar-Ness, "Bootstrap: A fast adaptive signal separator," in *ICASSP 92*, pp. II.325–II.328, 1992.

[7] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, , 1983.

[8] S. J. Nowlan and G. E. Hinton, "A soft decision-directed LMS algorithm for blind equalization," *IEEE Trans. on Communications*, vol. 41, pp. 275–279, feb. 1993.

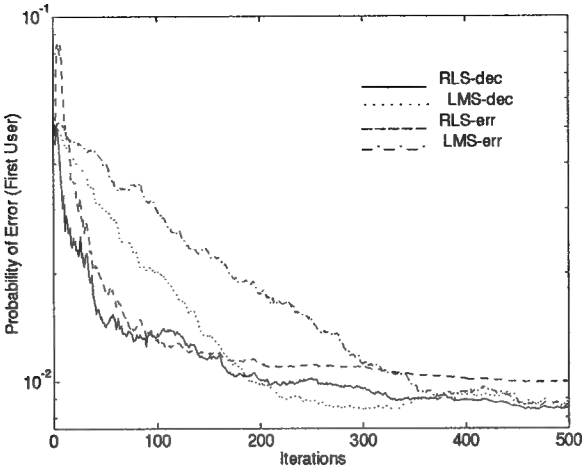


Figure 1: Learning curve of the probability of error of the first user (SNR=8dB,SSR=-10dB,N=4,ρ=0.15)

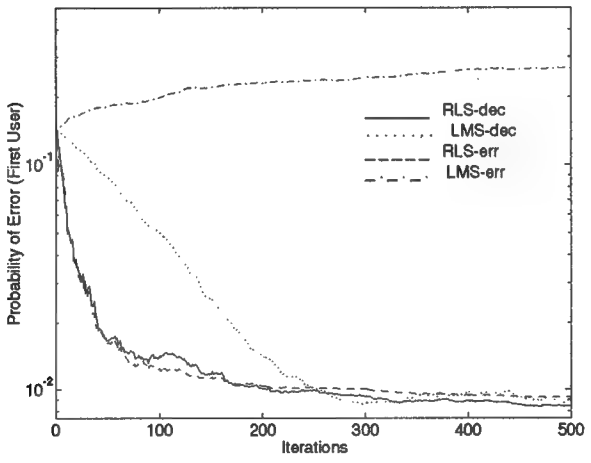


Figure 2: Learning curve of the probability of error of the first user (SNR=8dB,SSR=-5dB,N=4,ρ=0.15)

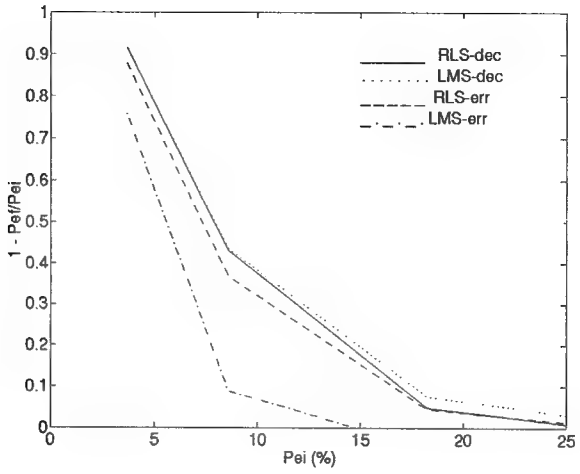


Figure 3: Convergence regions after 250 iterations for the probability of error of the first user (SNR=8dB,SSR=-10dB,N=4,ρ=0.15)

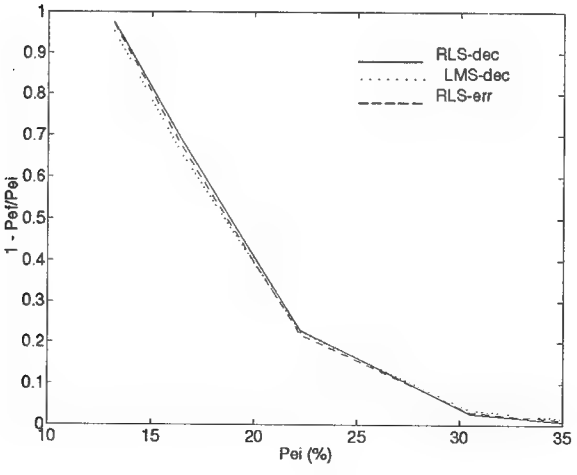


Figure 4: Convergence regions after 250 iterations for the probability of error of the first user (SNR=8dB,SSR=-5dB,N=4,ρ=0.15)

APPENDIX A: PART II

ADAPTIVE DETECTORS FOR MULTI-CHANNEL SIGNALS IN CODE DIVISION MULTIPLE ACCESS SYSTEMS

by
Alex Haimovich, Ruth Onn and Yeheskel Bar-Ness

ABSTRACT

This paper considers adaptive receivers for multi-channel cross-correlated signals. The specific application considered is the separation of signals in code division multiple access (CDMA) systems, but the methods developed are applicable to other similarly modeled problems. The minimum mean square error (MSE) and the signal decorrelation criteria are applied to derive feedforward and feedback detector structures. It is shown that in the noise-free case, and at steady state, these detectors indeed perform signal separation. The performance of these adaptive detectors in noisy environment is compared against each other and to the ideal decorrelator detector. The Feedback Minimum Error detector is found to have the closest performance to the ideal decorrelator.

I. INTRODUCTION

This paper considers adaptive receivers for code-division multiple-access (CDMA) signals. A number of low complexity structures have been proposed by Verdu [1], Aazhang [2], and others to solve the multi-user detection problem with little degradation vis-a-vis the optimal detector. In particular the decorrelating detector was shown to be near-far resistant, i.e., insensitive to the power of the interfering signals.

Most of the strategies for multi-user detection exploit the knowledge of the codes assigned to the users to decorrelate between the signals. Furthermore, it is customarily assumed that, following match filtering in each channel, the cross-correlations between the signals are known and can be eliminated by the fixed decorrelating receiver. Nevertheless, in practice, channel effects often cause unknown correlations at the output of the filters. This paper addresses the design of *adaptive* receivers for the separation of CDMA signals. The work has been inspired by structures previously suggested for blind signal separation and referred to as *bootstrap* algorithms [3].

In Section II we define the system model, review the single user and the fixed decorrelator detectors, and present the feedforward and feedback configurations for adaptive signal separation. Four adaptive detection criteria, two for the feedforward and two for the feedback configurations, are presented in Section III. LMS type algorithms are suggested for each detector. The performance of the adaptive detectors is compared by simulations in Section IV. Section V provides the conclusions.

II. SYSTEM MODEL

Consider an N-user synchronous CDMA system. We make the following basic assumptions:

(1) The sources are independent.

(2) The outputs of the channels consist of a linear mixture of the inputs.

(3) Number of inputs is equal to the number of sources.

The waveform (code) used by the n-th user is denoted by $g_n(t)$, $0 \leq t \leq T_s$, where T_s is the data bit duration. This waveform is used to modulate each of the n-th user information bits. The received waveform is given by $c_n(t) = g_n(t) * h_n(t)$, where $h_n(t)$ is the n-th channel (unknown) impulse response. The channel causes interchip interference, but it is assumed that the intersymbol interference is negligible. The received signal is applied to a bank of filters, which are matched to the signature waveforms $g_n(t)$. Due to the channel effects, the outputs of the match filters is represented by an unknown correlation matrix $\mathbf{A} = \{A_{ij}\}$. $\mathbf{A}$ is normalized such that $A_{nn} = 1$. The channel response is assumed not to vary with time; hence, $\mathbf{A}$ consists of fixed elements. It is also assumed that $\mathbf{A}$ is strictly diagonal dominant, i.e., $|A_{ij}| > \sum_{j \neq i} |A_{ij}|$. This assumption implies that the autocorrelation of the signal in any channel is larger than the sum of the cross-correlations between that channel and the other channels. With this assumption, $\mathbf{A}$ is invertible. Let $\mathbf{x}$ be a vector representing the bank of match filters outputs sampled at $t = iT_s$. Then $\mathbf{x}$ can be written:

$$\mathbf{x} = \mathbf{AEb} + \mathbf{z} \tag{1}$$

where $\mathbf{E}$ is the diagonal matrix, $\mathbf{E} = \text{diag}\left(\sqrt{\xi_1}, \ldots, \sqrt{\xi_N}\right)$, and $\mathbf{b}^T = (b_1, \ldots, b_N)$. The information symbols $b_n \in \{-1, 1\}$ are outcomes of identical, uniformly distributed, and independent between users random variables. ξ_n is the unknown energy of the received signal from user n. A time index is not used since decisions are based on observations in a single data bit interval. The noise vector $\mathbf{z}$, is assumed Gaussian with correlation matrix $\mathbf{R}(\mathbf{z}) = \sigma^2\mathbf{G}$, where $\mathbf{G}$ is the cross-correlation matrix at the output of the match filters when the input is white noise ($(\mathbf{G})_{ij} = \int_0^{T_s} g_i(t)g_j(t)\text{dt}$).

2.1 Single User Detector

The conventional single user detector makes its decision based on the sign of the appropriate element of $\mathbf{x}$. Thus the decision for the n-th user is given by:

$$\widehat{b}_n = \text{sgn}(x_n) \tag{2}$$

The n-th user error probability of the single user detector P_n^s can be calculated as follows:

$$\begin{aligned} P_n^s &= P\left[x_n > 0 \mid b_n = -1\right] \\ &= \sum_{\substack{b\in\{-1,1\}^N \\ b_n=-1}} P\left[x_n > 0 \mid \mathbf{b}\right] P\left[\mathbf{b} \mid b_n = -1\right] \\ &= 2^{1-N} \sum_{\substack{b\in\{-1,1\}^N \\ b_n=-1}} Q\left(\frac{\sqrt{\xi_n} - \sum_{j\neq n}\sqrt{\xi_j} A_{jn} b_j}{\sigma\left(\mathbf{G}\right)_{nn}}\right) \end{aligned} \tag{3}$$

where $P\left(\mathbf{b} \mid b_n = -1\right) = 2^{1-N}$.

2.2 Decorrelation Detector - Known Correlation Matrix A

If channel effects are ignored then $\mathbf{A} = \mathbf{G}$ is known. The data vector $\mathbf{b}$ can be recovered from the outputs of the bank of match filters as follows: $\mathbf{y} = \mathbf{G}^{-1}\mathbf{x} = \mathbf{G}^{-1}\left(\mathbf{GEb} + \mathbf{z}\right) = \mathbf{Eb} + \eta$. The gaussian noise vector η is zero mean and has covariance matrix $\mathbf{R}(\eta) = \mathbf{G}^{-1}\mathbf{R}(\mathbf{z})\mathbf{G}^{-T} = \sigma^2\mathbf{G}^{-T}$, $(\mathbf{G}^{-T} = (\mathbf{G}^{-1})^T)$. For the n-th user we have $y_n = \sqrt{\xi_n}b_n + \eta_n$, and $\widehat{b}_n = \mathrm{sgn}(y_n)$. Hence, the n-th user probability of error P_n^d is given by

$$P_n^d = Q\left(\frac{\sqrt{\xi_n}}{\sigma^2\left(\mathbf{G}^{-T}\right)_{nn}}\right) \tag{4}$$

In this work we are concerned with the design of multi-user detectors when the correlation matrix $\mathbf{A}$ is not known. In this context we distinguish between two basic configurations: *feedforward*, in which the adaptive filter operates on the inputs vector $\mathbf{x}$, and *feedback*, in which the adaptive filter operates on the outputs vector $\mathbf{y}$. Both configurations are shown schematically in Figure 1

2.3 Adaptive Detector - Feedforward Configuration

The n-th user output y_n, is obtained by applying the n-th channel filter coefficient vector $\mathbf{w}_n$ to the input vector $\mathbf{x}$. If the weight matrix for the forward configuration is defined $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_N]$, the output vector is given by $\mathbf{y} = \mathbf{WAEb} + \mathbf{Wz}$. The decision slicer output is given by $\widehat{b}_n = \mathrm{sgn}\left(y_n\right) = \mathrm{sgn}(\mathbf{w}_n^T\mathbf{x})$. For any specific weight vector $\mathbf{w}_n$, the probability of error for the n-th user can be calculated as follows:

$$\begin{aligned} P_n^f &= P\left[\widehat{b}_n > 0 \mid b_n = -1\right] \\ &= P\left[y_n > 0 \mid b_n = -1\right] \\ &= P\left[\mathbf{w}_n^T\mathbf{x} > 0 \mid b_n = -1\right] \\ &= 2^{1-N} \sum_{\substack{b\in\{-1,1\}^N \\ b_n=-1}} Q\left(\frac{t_{nn} - \sum_{j\neq n} t_{nj} b_j}{\sigma\sqrt{\mu_{nn}}}\right) \end{aligned} \tag{5}$$

where t_{nn} is the n-th element of the vector $\mathbf{t}_n = \mathbf{w}_n^T\mathbf{AE}$, and $\mu_{nn} = \mathbf{w}_n^T\mathbf{G}\mathbf{w}_n$ is zero-mean Gaussian noise with variance $\sigma^2\mu_{nn}$.

2.4 Adaptive Detector - Feedback Configuration

With the feedback adaptive multi-user detector the output y_n results from subtracting a linear combination of the other outputs $\mathbf{v}_n^T\tilde{\mathbf{y}}_n$, from the n-th user input x_n, $y_n = x_n - \mathbf{v}_n^T\tilde{\mathbf{y}}_n$, where $\mathbf{v}_n$ denotes the $N \times 1$ feedback weight vector. $\tilde{\mathbf{y}}_n$ is the vector of outputs obtained from $\mathbf{y}$ by setting $y_n = 0$. The last relation can be rewritten $x_n = y_n + \mathbf{v}_n^T\tilde{\mathbf{y}}_n = \mathbf{v}_n^T\mathbf{y}$, where $v_{nn} = 1$. We define the weight matrix for the feedback configuration $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N]^T$, which relates the output of the adaptive receiver to its input, $\mathbf{x} = \mathbf{V}\mathbf{y}$. If $\mathbf{V}$ is invertible then the output $\mathbf{y}$ can be obtained as $\mathbf{y} = \mathbf{V}^{-1}\mathbf{x}$. Otherwise, the generalized inverse can be used to compute $\mathbf{y}$. The probability of error can then be found from eq. (5) using $\mathbf{V}^{-1}$ in place of $\mathbf{W}$.

III. Signal Decorrelation Criteria

In this section we define four signal separation criteria. Two are based on the minimization of the mean square error (MSE) between the channel output and a reference supplied by the detected output. The other two criteria are for decorrelation of the outputs.

3.1 Feedforward Minimum Error Detector

The *Feedforward Minimum Error* (FME) detector minimizes the mean error between its output and a reference signal. Typically, the reference is initially supplied by a training signal. When the adaptive weights converged and the errors with respect to the training signal are small, the detector is switched to operate in decision directed mode, with the reference signal supplied by the detected symbol. The FME detector realizes the optimal linear receiver for each of the users. The filter coefficients may be updated using the LMS algorithm:

$$\mathbf{w}_n(k+1) = \mathbf{w}_n(k) - \mu e(k)\,\mathbf{x}(k) \tag{6}$$

where $e(k) = \widehat{b}_n(k) - y_n(k)$ and $y_n(k) = \mathbf{w}_n^T(k)\mathbf{x}(k-1)$. The steady state solution of the FME detector in terms of the weight matrix is given by,

$$\begin{aligned} \mathbf{W}_{\text{FME}} &= \arg\min_{W} \mathrm{E}\left[\left\|\widehat{\mathbf{b}} - \mathbf{W}^T\mathbf{x}\right\|^2\right] \\ &= \mathbf{R}_x^{-1}\mathbf{R}_{x\widehat{b}} \end{aligned} \tag{7}$$

where $\mathbf{R}_x = \mathrm{E}\left[\mathbf{x}\mathbf{x}^T\right]$ is the input correlation matrix, and $\mathbf{R}_{x\widehat{b}} = \mathrm{E}\left[\mathbf{x}\widehat{\mathbf{b}}^T\right]$ is the cross-correlation matrix between the input and the estimated symbol. This criterion leads to signal separation in the noise-free case and to power-inversion interference cancellation when additive noise is present. In the noise-free case the input correlation matrix can be written:

$$\begin{aligned} \mathbf{R}_x &= \mathbf{AE}\mathrm{E}\left[\mathbf{b}\mathbf{b}^T\right]\mathbf{E}\mathbf{A}^T \\ &= \mathbf{A}\mathbf{E}^2\mathbf{A}^T \end{aligned} \tag{8}$$

where $\mathbf{E}^2$, is a diagonal matrix of the signal energies, and we used $\mathrm{E}\left[\mathbf{b}\mathbf{b}^T\right] = \mathbf{I}$, since the signals are assumed independent between users.
The cross-correlation matrix is:

$$\begin{aligned} \mathbf{R}_{x\widehat{b}} &= \mathrm{E}\left[\mathbf{A}\mathbf{E}\mathbf{b}\widehat{\mathbf{b}}^T\right] \\ &= p\mathbf{A}\mathbf{E} \end{aligned} \tag{9}$$

where we used, $p = 1 - 2Pr\left[\widehat{b}_j \neq b_j\right] = 1 - 2P_e$, and the result [4]:

$$\mathrm{E}\left[b_i\widehat{b}_j\right] = \begin{cases} 1 - 2P_e & i = j \\ 0 & i \neq j \end{cases} \tag{10}$$

Then the weights are given by $\mathbf{W}_{\mathrm{FME}} = p\mathbf{A}^{-T}\mathbf{E}^{-1}$. In the absence of noise, this linear transformation applied to the input vector, recovers the transmitted signals, $\mathbf{y} = \mathbf{W}_{\mathrm{FME}}^T\mathbf{x} = p\mathbf{b}$.

3.2 Feedforward Zero Correlation Detector

The *Feedforward Zero Correlation Detector* (FZC) is different from the FME detector. The FME weight vector for any user is controlled by the error derived solely from the slicer output corresponding to that user. In contrast, the n-th channel FZC weight vector is controlled by *all* other slicers' outputs. Separation is achieved through decorrelation of the n-th channel slicer input from all the other slicers' outputs. LMS zero correlation algorithms have been proven to have faster learning curves than LMS minimum error algorithms [4]. The LMS zero correlation algorithm for the n-th channel is given by:

$$\mathbf{w}_n(k+1) = \mathbf{w}_n(k) - \mu y_n(k)\,\tilde{\mathbf{b}}_n(k) \tag{11}$$

where $\tilde{\mathbf{b}}_n^T = \left[\widehat{b}_1, \cdots, \overset{n}{0}, \cdots, \widehat{b}_N\right]$ is obtained from the vector of estimated outputs $\widehat{\mathbf{b}}$, by setting $\widehat{b}_n = 0$. The input to the n-th channel is passed with unity gain, $w_{nn} = 1$. The algorithm converges in the mean when $\mathrm{E}\left[y_n\tilde{\mathbf{b}}_n\right] = 0$, or equivalently, when $\mathrm{E}\left[y_n\widehat{\mathbf{b}}\right] = \alpha_n\mathbf{u}_n$, where $\mathbf{u}_n$ is a unit vector with the n-th element set to 1, and α_n is chosen such that $w_{nn} = 1$. Using matrix notation, decorrelation is achieved for $\mathrm{E}\left[\mathbf{y}\widehat{\mathbf{b}}^T\right] = \mathrm{diag}\left(\alpha_1, \ldots, \alpha_N\right) \equiv \mathbf{D}$. In terms of the weight matrix $\mathbf{W}$, and using eq. (9), this condition can be written:

$$\begin{aligned} \mathrm{E}\left[\mathbf{y}\widehat{\mathbf{b}}^T\right] &= \mathrm{E}\left[\mathbf{W}_{\mathrm{FZC}}^T(\mathbf{A}\mathbf{E}\mathbf{b} + \mathbf{z})\widehat{\mathbf{b}}^T\right] \\ &= p\mathbf{W}_{\mathrm{FZC}}^T\mathbf{A}\mathbf{E} \end{aligned} \tag{12}$$

Hence,

$$\mathbf{W}_{\mathrm{FZC}} = \mathbf{D}\left(p\mathbf{A}\mathbf{E}\right)^{-1} \tag{13}$$

Indeed, when this weight matrix is applied to the input $\mathbf{x}$, each channel output consists of a scaled output of that channel's user and noise: $\mathbf{y} = \mathbf{W}_{\mathrm{FZC}}^T\mathbf{x} = \mathbf{W}_{\mathrm{FZC}}^T\left(\mathbf{A}\mathbf{E}\mathbf{b} + \mathbf{z}\right) = p^{-1}\mathbf{D}\mathbf{b} + \mathbf{W}_{\mathrm{FZC}}^T\mathbf{z}$. Notice that with the decorrelation criterion it is not necessary to assume a noise-free environment in order to achieve separation.

3.3 Feedback Minimum Error Detector

The *Feedback Minimum Error* (BME) detector minimizes the MSE between the input to the n-th channel slicer y_n, and a reference signal supplied by the slicer output $\hat{b}_n$. It achieves that by subtracting from x_n a linear combination of the other users' outputs, $\mathbf{v}_n^T\,\tilde{\mathbf{y}}_n$. The error is given by $e(k) = \hat{b}_n(k) - y_n(k)$, and the MSE is defined by $\varepsilon = \mathrm{E}[e^2(k)]$. Using $\mathbf{y} = \mathbf{V}^{-T}\mathbf{x}$, the BME weight matrix is given by:

$$\begin{aligned}\mathbf{V}_{\mathrm{BME}} &= \arg\min_{V^{-1}} \mathrm{E}\left[\left\|\hat{\mathbf{b}} - \mathbf{V}^{-T}\mathbf{x}\right\|^2\right] \\ &= \mathbf{R}_{x\hat{b}}^{-1}\mathbf{R}_x\end{aligned}$$

Using eqs. (8) and (9) we get for the noise-free case, $\mathbf{V}_{\mathrm{BME}} = (p^{-1}\mathbf{E}^{-1}\mathbf{A}^{-1})\left(\mathbf{A}\mathbf{E}^2\mathbf{A}^T\right) = p^{-1}\mathbf{E}\mathbf{A}^T$. The feedback filter coefficients $\mathbf{V}_{\mathrm{BME}}$ achieve signal separation: $\mathbf{y} = \mathbf{V}_{\mathrm{BME}}^{-T}\mathbf{x} = p\mathbf{b}$. The n-th channel BME feedback filter coefficients are updated using the following LMS algorithm:

$$\mathbf{v}_n(k+1) = \mathbf{v}_n(k) + \mu e(k)\,\tilde{\mathbf{y}}_n(k) \tag{14}$$

With the n-th element of $\tilde{\mathbf{y}}_n$ set to zero, the algorithm does not control v_{nn}, which is set to 1. Convergence in the mean is achieved when the error between the input to the slicer and its output is decorrelated from the inputs to the other slicers.

3.4 Feedback Zero Correlation Detector

The *Feedback Zero Correlation* (BZC) detector is the feedback configuration analogous to the FZC detector. The updating of the n-th channel feedback weight vector is given by the following LMS decorrelation algorithm:

$$\mathbf{v}_n(k+1) = \mathbf{v}_n(k) - \mu y_n(k)\,\tilde{\mathbf{b}}_n(k) \tag{15}$$

where the vector $\tilde{\mathbf{b}}_n$ has been defined previously. The algorithm does not control v_{nn}, which is set to 1. This algorithm converges in the mean when $\mathrm{E}[y_n\hat{\mathbf{y}}_n] = 0$, or equivalently, when $\mathrm{E}\left[\mathbf{y}\hat{\mathbf{b}}^T\right] = \mathrm{diag}(\alpha_1, \ldots, \alpha_N) \equiv \mathbf{D}$, where the elements of $\mathbf{D}$ are such that $v_{nn}(k) = 1$. The feedback filter coefficients $\mathbf{V}_{\mathrm{BZC}}$ can be found as follows:

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}\hat{\mathbf{b}}^T\right] &= \mathrm{E}\left[\mathbf{V}_{\mathrm{BZC}}^{-T}(\mathbf{A}\mathbf{E}\mathbf{b} + \mathbf{z})\,\hat{\mathbf{b}}^T\right] \\ &= p\mathbf{V}_{\mathrm{BZC}}^{-T}\mathbf{A}\mathbf{E}\end{aligned} \tag{16}$$

Hence,

$$\mathbf{V}_{\mathrm{BZC}} = p\mathbf{D}^{-1}\mathbf{E}\mathbf{A}^T \tag{17}$$

Indeed, when this weight matrix is applied to the input $\mathbf{x}$, each channel output consists of the scaled corresponding user signal and noise: $\mathbf{y} = \mathbf{V}_{\mathrm{BZC}}^{-T}\mathbf{x} = \mathbf{V}_{\mathrm{BZC}}^{-T}(\mathbf{A}\mathbf{E}\mathbf{b} + \mathbf{z}) = p\mathbf{D}^{-1}\mathbf{b} + \mathbf{V}_{\mathrm{BZC}}^{-T}\mathbf{z}$. Similar to the FZC detector case, the noise-free assumption is not required by the BZC detector for signal separation.

IV. SIMULATIONS RESULTS

A performance comparison of the operation of those detectors in noise has been carried out by simulations. The simulation model consisted of a $N = 4$ channel system, with fixed cross-correlations of 0.2. Probabilities of error were calculated after convergence of the four algorithms in eqs. (6), (11), (14), and (15), respectively. Figure 2 shows the probability of error of the first signal as a function of the power ratio of the second to the first signal, with the powers for signals 3 and 4 kept constant, and SNR = 8 dB. Curves for probability of error for decorrelation with known cross-correlation matrix **A** and single user detector are also shown. It can be observed that the performance of the adaptive BME detector is the best among the methods considered. It is actually slightly better even than the decorrelator detector with known **A**, for the whole range of power of the second user signal. The advantage of the BME method is the consequence of minimizing an error, rather than imposing the zero correlation condition. It is also noted that the FME detector is sensitive to the power level of the interference, and has the poorest performance among the adaptive detectors when the interference is lower than the desired signal. Figure 3 shows the probability of error of the first user as a function of the SNR, when the power ratio of the second user to the first user is 5 dB. The BME has the closest performance to the known correlation matrix detector. The FZC and BZC detectors are slightly worse.

V. CONCLUSIONS

Four adaptive methods for signal separation are compared. Two methods are based on a feedforward configuration, the feedforward minimum error and the feedforward zero correlation, and two methods are based on a feedback configuration, the feedback minimum error and the feedback zero correlation. The feedback minimum error is shown to have the best performance in terms of probability of error.

References

[1] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Information Theory*, vol. IT-35, pp. 123–136, Jan. 1989.

[2] M. K. Varanasi and B. Aazhang, "Near-optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Communications*, vol. COM-39, pp. 725–736, May 1991.

[3] A. Dinc and Y. Bar-Ness, "Bootstrap: A fast adaptive signal separator," in *ICASSP '92*, pp. II.325–II.328, 1992.

[4] A. Haimovich, R. Manzo, and Y. Bar-Ness, "Fast decorrelating algorithms for signal separation," Submitted to Globecom '94.

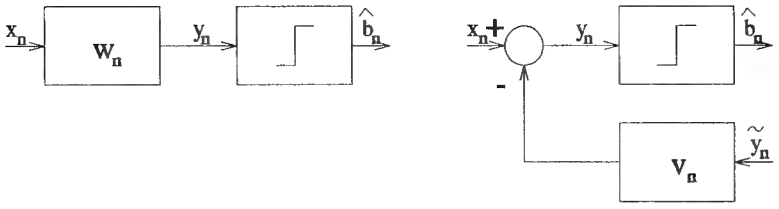


Figure 1: Feedforward and feedback configurations of the adaptive multiuser detector.

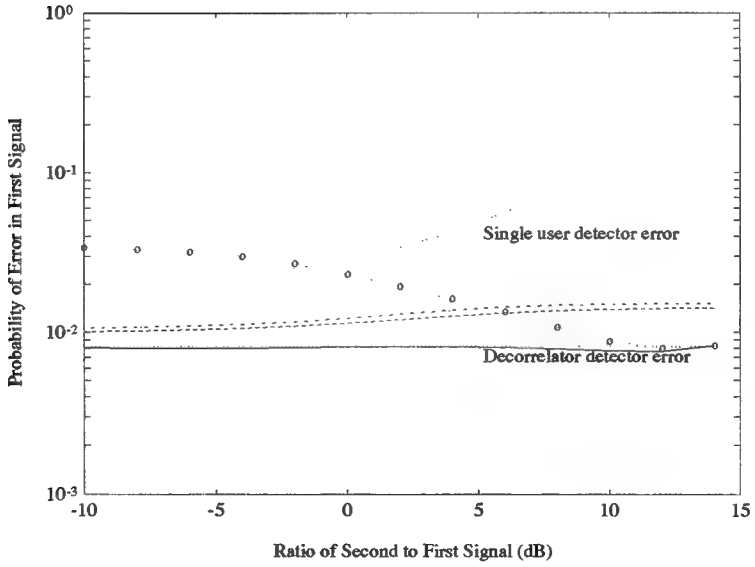


Figure 2: Probability of error in first signal versus the ratio of the second to first signal. The methods compared are: FZC (dashed line), BZC (dash-dot), FME (regular LMS) (circles on dots), and BME (continuous line). Also on the plot are the lines indicating the single user detector, and the de-c orrelator detector probabilities of error (dotted lines).

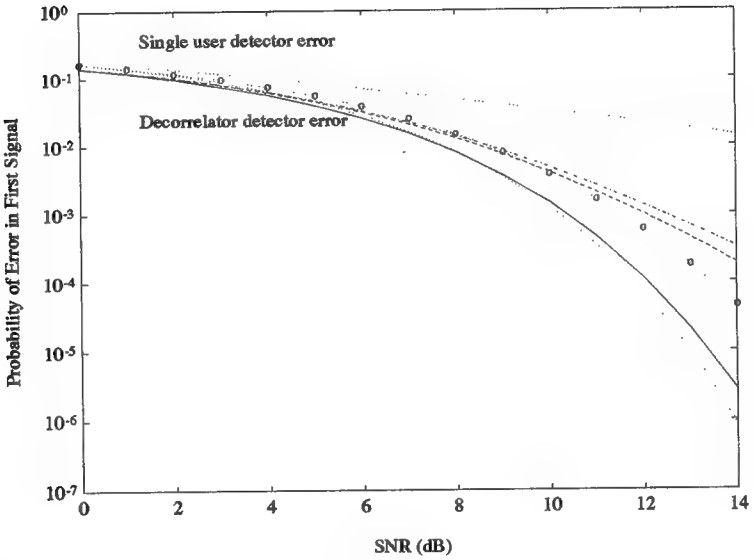


Figure 3: Probability of error versus SNR for first signal. The methods compared are: FZC (dashed line), BZC (dash-dot), FME (regular LMS) (circles on dots), and BME (continuous line). Also on the plot are the lines indicating the single user detector, and the de-c orrelator detector probabilities of error (dotted lines).

APPENDIX B: PART I

ERROR PERFORMANCE OF SYNCHRONOUS MULTIUSER CODE DIVISION MULTIPLE ACCESS DETECTOR WITH MULTIDIMENSIONAL ADAPTIVE CANCELER

by
Zoran Siveski, Yeheskel Bar-Ness and David W. Chen

ABSTRACT

A code division multiple access detector that employs a combination of a decorrelator and a multidimensional interference canceler is considered. The weights of the canceler are adaptively controlled using a steepest descent algorithm. The probability of error is evaluated and compared to those of the decorrelating detector and a similar, fixed-weights scheme that requires an estimation of received signal energies. It is shown that the proposed two-stage detector provides substantial improvement over the decorrelating detector, particularly in the presence of strong interfering signals. This is especially noticeable in the high bandwidth efficiency cases. It is also observed that in the presence of weak interferers, the proposed scheme performs better than its fixed-weights counterpart.

I. INTRODUCTION

A conventional single-user detector implemented in the Code Division Multiple Access (CDMA) system consists of a bank of filters, each one matched to the signature sequence of the particular user. The sampled output of each matched filter, besides the desired signal, contains the residual interference from all other users. The presence of a strong interference often makes it impossible to detect the weak user, a condition referred to as a near-far problem. In [1] a receiver that is optimum in the multiuser interference environment was proposed and shown to eliminate the near-far problem and provide much improved performance. The improvement comes at the expense of high computational complexity. A class of suboptimum receivers that uses decorrelating detectors and which is based on the linear transformation of the sampled matched filters' outputs was considered in [2] and [3]. The decorrelating decision-feedback detector presented in [4] utilizes the difference in received users' energies where the decisions of the stronger users are used to eliminate interference on weaker users. Another approach for suboptimum multi-user detectors with low complexity was proposed in [5] and [6], where in order to perform detection of the desired user, tentative decisions on information bits of all other users are made. The estimate of the multiple access interference is then obtained and is subtracted from the desired signal. The performance of some of these suboptimum schemes is close to the performance of the optimum detector, particularly when the power of the interferers increases, they become indistiguishable. However, these schemes have to perform an estimation of the received signal energies, knowledge

of which is required for the detectors' proper operation.
The detector for synchronous CDMA systems proposed here consists of a decorrelator and a nonlinear multidimensional interference canceler whose weights adaptively adjust to the incoming signal. In Section 2, the system model and the derivation of the expressions for the steady state values of the weights and the output error probability are presented. Numerical examples and a discussion are presented in Section 3.

II. DETECTOR: MODEL AND ANALYSIS

A synchronous CDMA receiver considered here consists of a bank of filters (matched to the users' signature sequences) front-end, followed by samplers and the decision system, as depicted in Fig. 1.

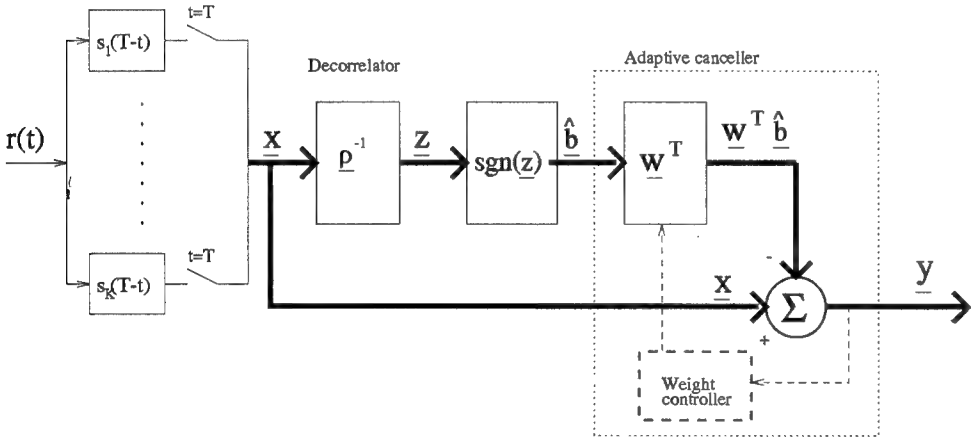


Figure 1: Proposed receiver scheme

The received signal $r(t)$ is expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i) \sqrt{a_k} s_k(t - iT) + n(t) \tag{1}$$

where $b_k(i) \in \{-1, +1\}$ is the k-th user's data bit in the i-th time interval and $n(t)$ is the additive, zero mean, white Gaussian noise with the two-sided power spectral density $N_0/2$. The received energy of the k-th user signal, unknown to the receiver, is denoted as a_k. While in reality it is slowly time varying, it is assumed to remain unchanged over the transmission horizon of each user. The signature sequence $s_k(t)$, of the same duration T as a data bit, is known to the receiver. The sampled outputs of the bank of matched filters in the i-th bit interval can be expressed as:

$$\boldsymbol{x}(i) = \boldsymbol{\mathcal{P}} \boldsymbol{A} \boldsymbol{b}(i) + \boldsymbol{n}(i) \tag{2}$$

For the sake of convenience the index i will be omitted elsewhere in the text. In (2), $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]^T$, $\boldsymbol{b} = [b_1, b_2, \ldots, b_K]^T$, $\boldsymbol{A} = \text{diag}[\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_K}]$ is a diagonal

matrix and $\boldsymbol{n} = [n_1, n_2, \ldots, n_K]^T$. The (k,j)th element of the symmetric crosscorrelation matrix $\boldsymbol{\mathcal{P}}$ is defined as:

$$\rho_{kj} = \int_0^T s_k(t) s_j(t) dt \qquad k, j \in (1, 2, \ldots, K)$$

with $\rho_{kk} = 1$. The covariance matrix of a zero mean Gaussian noise vector $\boldsymbol{n}$ is,

$$E\{\boldsymbol{n}\boldsymbol{n}^T\} = \frac{N_0}{2}\boldsymbol{\mathcal{P}}$$

The outputs of the decorrelator are given as:

$$\boldsymbol{z} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\Gamma}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\xi} \tag{3}$$

where $\boldsymbol{\Gamma}$ is the inverse of $\boldsymbol{\mathcal{P}}$, and $\boldsymbol{\xi} = \boldsymbol{\Gamma}\boldsymbol{n}$. Let $\hat{\boldsymbol{b}} = \left[\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_K\right]^T$ be the decision output of the decorrelator defined as:

$$\hat{\boldsymbol{b}} = \operatorname{sgn}(\boldsymbol{z}) \tag{4}$$

Then the canceler's output is given by:

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{W}^T\hat{\boldsymbol{b}} \tag{5}$$

where

$$\boldsymbol{W} = \begin{bmatrix} 0 & w_{12} & \ldots & w_{1K} \\ w_{21} & 0 & \ldots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \ldots & 0 \end{bmatrix}$$

The output for the k-th user can be expressed as:

$$y_k = x_k - \boldsymbol{w}_k{}^T\hat{\boldsymbol{b}}_k \tag{6}$$

where $\boldsymbol{w}_k$ is the k-th column vector of $\boldsymbol{W}$ with the element w_{kk} deleted, and $\hat{\boldsymbol{b}}_k$ is vector obtained from $\hat{\boldsymbol{b}}$ by deleting the element $\hat{b}_k$. The multidimensional canceler structure is the same as the one termed in array processing sidelobe canceler, except for the fact that for auxiliary signals we are using the decisions outputs of the decorrelator. For controling the weights we use steepest descent algorithm which simultaneously minimizes the output signal powers $E\{y_k^2\}$. That is, for the k-th output, the optimum weights are obtained by iterative search:

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k - \frac{\mu}{2}\frac{\partial}{\partial \boldsymbol{w}_k}E\{y_k^2\} = \boldsymbol{w}_k - \frac{\mu}{2}\frac{\partial}{\partial \boldsymbol{w}_k}E\{(x_k - \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k)^2\} = \boldsymbol{w}_k + \mu E\{y_k\hat{\boldsymbol{b}}_k\} \tag{7}$$

Clearly such an algorithm, by minimizing the output power, forces the correlation between the desired output signal and vector of tentative decisions of interfering signals to zero. In practice, the expactation operation is replaced by time averaging.

The steady state values can be evaluated as follows:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{w}_k} E\{y_k^2\} = 0 &= \frac{\partial}{\partial \boldsymbol{w}_k} E\{(x_k - \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k)(x_k - \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k)^T\} \\ &= \frac{\partial}{\partial \boldsymbol{w}_k} E\{x_k^2 - 2x_k \hat{\boldsymbol{b}}_k^T \boldsymbol{w}_k + \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T \boldsymbol{w}_k\} \\ &= -2E\{x_k \hat{\boldsymbol{b}}_k\} + 2E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\} \boldsymbol{w}_k \end{aligned} \tag{8}$$

From the definition of $\hat{\boldsymbol{b}}$ it is obvious that $E\{b_i \hat{b}_j\} = 0$ for $i \neq j$. Also, it should be noted that $E\{\boldsymbol{\xi} \boldsymbol{n}^T\} = (N_0/2)\boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix. Then, it is easily shown that:

$$E\{x_k \hat{\boldsymbol{b}}_k\} = E\left\{\left[\sqrt{a_k} b_k + \boldsymbol{\rho}_k^T \boldsymbol{A}_k \boldsymbol{b}_k + n_k\right] \hat{\boldsymbol{b}}_k\right\} = \left[\boldsymbol{A}_k E\{\boldsymbol{b}_k \hat{\boldsymbol{b}}_k^T\}\right]^T \boldsymbol{\rho}_k \tag{9}$$

where $\boldsymbol{A}_k$ is a diagonal $(K-1) \times (K-1)$ submatrix of $\boldsymbol{A}$ with its k-th diagonal entry removed, $\boldsymbol{\rho}_k$ is a $(K-1) \times 1$ k-th column vector obtained from $\boldsymbol{\mathcal{P}}$ by deleting the element ρ_{kk}, and $\boldsymbol{b}_k$ is a $(K-1) \times 1$ vector obtained from $\boldsymbol{b}$ by deleting the element b_k. Clearly $E\{\boldsymbol{b}_k \hat{\boldsymbol{b}}_k^T\}$ is diagonal, therefore the system of $K-1$ linear equations (8), together with (9), gives the steady state values of the weights affecting the k-th output as:

$$\boldsymbol{w}_k = \left[E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\}\right]^{-1} \boldsymbol{A}_k E\{\boldsymbol{b}_k \hat{\boldsymbol{b}}_k^T\} \boldsymbol{\rho}_k \tag{10}$$

The expectations in the above expression can be computed as follows:

$$\begin{aligned} E\{\hat{b}_i \hat{b}_j\} &= E\left\{\operatorname{sgn}\left(\sqrt{a_i} b_i + \xi_i\right) \operatorname{sgn}\left(\sqrt{a_j} b_j + \xi_j\right)\right\} \\ &= \frac{1}{2} \sum_{b_i, b_j} \left(Pr\{\xi_i < \sqrt{a_i} b_i, \xi_j < \sqrt{a_j} b_j\} - Pr\{\xi_i < \sqrt{a_i} b_i, \xi_j > \sqrt{a_j} b_j\}\right) \end{aligned}$$

where each probability term in the above sumation defines four integrals. That is:

$$E\{\hat{b}_i \hat{b}_j\} = \frac{1}{2} \sum_{l=1}^{4} \int\int_{D_l} f_{\xi_i \xi_j} d\xi_i d\xi_j - \frac{1}{2} \sum_{l=5}^{8} \int\int_{D_l} f_{\xi_i \xi_j} d\xi_i d\xi_j \qquad i \neq j \tag{11}$$

where $f_{\xi_i \xi_j}$ denotes the bivariate Gaussian density function, and D_l an appropriate rectangular region of integration. The covariance matrix of the two, zero mean, random variables ξ_i and ξ_j can be obtained as a submatrix of the $K \times K$ covariance matrix

$$E\{\boldsymbol{\xi} \boldsymbol{\xi}^T\} = E\{(\boldsymbol{\Gamma} \boldsymbol{n})(\boldsymbol{\Gamma} \boldsymbol{n})^T\} = \frac{N_0}{2} \boldsymbol{\Gamma} \tag{12}$$

Also:

$$E\{b_i \hat{b}_i\} = 1 - 2Pr\{\hat{b}_i \text{ in error}\} = 1 - 2Q\left(\sqrt{\frac{a_i}{\gamma_{ii} N_0/2}}\right) \tag{13}$$

where:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$$

Let the decision output of the second stage $\hat{\hat{\boldsymbol{b}}} = \left[\hat{\hat{b}}_1, \hat{\hat{b}}_2, \ldots, \hat{\hat{b}}_K\right]^T$ be defined as:

$$\hat{\hat{\boldsymbol{b}}} = \text{sgn}(\boldsymbol{y}) \tag{14}$$

The output error probability for k-th user is evaluated as follows:

$$\begin{aligned} P_{e_k} &= E_{b_k, \boldsymbol{b}_k, \hat{\boldsymbol{b}}_k} Pr\{\hat{\hat{b}}_k \text{ in error} | b_k, \boldsymbol{b}_k, \hat{\boldsymbol{b}}_k\} \\ &= \frac{1}{2} \sum_{\boldsymbol{b}_k, \hat{\boldsymbol{b}}_k} \Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\} \\ &+ Pr\{n_k < -\sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\} \Big] Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} Pr\{\boldsymbol{b}_k\} \end{aligned}$$

Since $Pr\{\boldsymbol{b}_k\} = 2^{-(K-1)}$ and n_k is a zero mean Gaussian random variable we can write:

$$\begin{aligned} P_{e_k} &= 2^{-K} \sum_{\boldsymbol{b}_k, \hat{\boldsymbol{b}}_k} \Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\} \\ &+ Pr\{n_k > \sqrt{a_k} + \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k - \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\} \Big] Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} \end{aligned} \tag{15}$$

It can be shown (see Appendix) that $Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} = Pr\{-\hat{\boldsymbol{b}}_k | -\boldsymbol{b}_k\}$, and therefore:

$$P_{e_k} = 2^{-(K-1)} \sum_{\boldsymbol{b}_k, \hat{\boldsymbol{b}}_k} \Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\} \Big] Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} \tag{16}$$

Finally Substituting (10) into (16) and recognizing that $\boldsymbol{A}_k$ is diagonal, we obtain for the k-th user error probability:

$$P_{e_k} = 2^{-(K-1)} \sum_{\boldsymbol{b}_k, \hat{\boldsymbol{b}}_k} Q\left(\frac{\sqrt{a_k} - \left(\boldsymbol{b}_k^T - \hat{\boldsymbol{b}}_k^T \left[E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\} \right]^{-1} E\{\boldsymbol{b}_k \hat{\boldsymbol{b}}_k^T\} \right) \boldsymbol{A}_k \boldsymbol{\rho}_k}{\sqrt{N_0/2}} \right) Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} \tag{17}$$

$Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\}$ is the integral of the $(K-1)$-variate Gaussian density function

$$Pr\{\hat{\boldsymbol{b}}_k | \boldsymbol{b}_k\} = \frac{1}{\sqrt{(2\pi)^{K-1} |\boldsymbol{\Xi}_k|}} \int_{\substack{\hat{b}_i \xi_i = -\sqrt{a_i} b_i \hat{b}_i \\ i=1,\ldots,K \quad i \neq k}}^{\infty} \cdots \int exp\left(-\frac{1}{2} \boldsymbol{\xi}_k^T \boldsymbol{\Xi}_k^{-1} \boldsymbol{\xi}_k \right) d\boldsymbol{\xi}_k$$

where $\boldsymbol{\xi}_k$ is a vector obtained from $\boldsymbol{\xi}$ by deleting the element ξ_k and $\boldsymbol{\Xi}_k$ is the $(K-1) \times (K-1)$ covariance matrix of $\boldsymbol{\xi}_k$ obtained from (12) by eliminating the k-th row and column. Introducing $\boldsymbol{\lambda}_k = \boldsymbol{A}_k^{-1} \boldsymbol{\xi}_k$, $\boldsymbol{\Gamma}_k = (N_0/2)^{-1} \boldsymbol{\Xi}_k$, and $\boldsymbol{R}_k = \text{diag}\left[\sqrt{SNR_1}, \ \ldots \ ,\right.$

$\sqrt{SNR_{k-1}}, \sqrt{SNR_{k+1}}, \ldots, \sqrt{SNR_K}\Big]$, where $SNR_i = a_i/N_0$, the above integral becomes:

$$Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} = \frac{\prod_{\substack{i=1\\ i\neq k}}^{K}\sqrt{SNR_i}}{\sqrt{\pi^{K-1}|\boldsymbol{\Gamma}_k|}} \int_{\substack{\hat{b}_i\lambda_i=-b_i\hat{b}_i\\ i=1,\ldots,K\\ i\neq k}}^{\infty}\cdots\int exp\left(-\boldsymbol{\lambda}_k^T\boldsymbol{R}_k\boldsymbol{\Gamma}_k^{-1}\boldsymbol{R}_k\boldsymbol{\lambda}_k\right)d\boldsymbol{\lambda}_k$$

The probability of error can be expressed as:

$$P_{e_k} = 2^{-(K-1)}\sum_{\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k} Q\left[\sqrt{2}\left(\sqrt{SNR_k} - \left(\boldsymbol{b}_k^T - \hat{\boldsymbol{b}}_k^T\left[E\{\hat{\boldsymbol{b}}_k\hat{\boldsymbol{b}}_k^T\}\right]^{-1}E\{\boldsymbol{b}_k\hat{\boldsymbol{b}}_k^T\}\right)\boldsymbol{R}_k\boldsymbol{\rho}_k\right)\right] Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} \tag{18}$$

III. NUMERICAL EXAMPLES AND DISCUSSION

In this section, two sets of numerical examples for the bit error probabilities of the proposed adaptive detector are calculated and presented. For comparison purposes, the error performances of the corresponding fixed-weights detector of Ref. [6] (in which the weights are set, based on the knowledge of the received signals' energies, to $w_{ji} = \rho_{ij}\sqrt{a_j}$), and those of the decorrelating detector are included.

The first example depicted in Fig. 2 is a simple, two-user case, but it nevertheless provides some insight into the steady state behavior of the adaptive detector. The crosscorrelation coefficient taken to be 0.7, can certainly be considered as representing a high bandwidth efficiency case. The SNR of user 1 is set to 8 dB, while the SNR of user 2, relative to user 1, varies from -10 to 8 dB. From this figure we notice first that when the interferer (user 2) is strong, user 1 achieves the performance of the single user system. Also noticeable is the fact that the adaptive system performs the same as the system with fixed weights [6]. On the other hand, with a weak interferer, and consequently, poor tentative decision $\hat{b}_2$, the adaptive system slightly outperforms the corresponding fixed-weights scheme. The improvement can be explained by comparing their respective amounts of residual interference. In the former, from (6), it is equal to $\rho_{12}\sqrt{a_2}\left[b_2 - E\{b_2\hat{b}_2\}\hat{b}_2\right]$, which is certainly less than in the latter case of $\rho_{12}\sqrt{a_2}\left[b_2 - \hat{b}_2\right]$. For almost all SNR's the performance of the decorrelating detector compares unfavorably to the other two.

In the second set of examples, Gold sequences of length seven (frequently used in the literature, e.g., [4] and [6]) are chosen for signature waveforms. The rationale for such a choice is that Gold sequences are regularly used in an asynchronous CDMA environment and the study of its proposed synchronous counterpart may provide a useful indication of the performance of the former. The crosscorrelation matrix $\boldsymbol{\mathcal{P}}$ in this case is:

$$\boldsymbol{\mathcal{P}} = \frac{1}{7}\begin{bmatrix} 7 & -1 & 3 & 3 & 3 \\ -1 & 7 & -1 & 3 & -1 \\ 3 & -1 & 7 & -1 & -1 \\ 3 & 3 & -1 & 7 & -1 \\ 3 & -1 & -1 & -1 & 7 \end{bmatrix}$$

The two-user case with Gold sequences is shown in Fig. 3. As expected, in this low bandwidth efficiency scenario ($\rho_{12} = -1/7$) the decorrelating detector performs as well as the other two schemes. By adding an additional user, Fig. 4, the decorrelating detector begins to exibit its inadequacy. The adaptive and the fixed-weights scheme show virtually identical performance, with the former being only slightly better for weak interferers. With the number of simultaneous users increasing further to K=4 and K=5, as in Figs. 5 and 6, respectively (note the change of the vertical axis scaling), certain trends become more obvious. Due to its unacceptable high probability of error, the decorrelating detector clearly does not represent an appropriate choice. When the interferers are strong, both the adaptive and the fixed-weights scheme achieve the performance of the single user. The former provides better error performance with weak interferers.
Finally, Fig. 7 depicts the error probabilities in the case when all users maintain the same SNR's. It shows the extent of the performance deterioration of the system as the number of simultaneous users increases.

IV. CONCLUSION

The adaptive code division multiple access detector proposed here was shown to provide significantly better steady state error performance than the decorrelating detector. This is especially true under the most critical conditions for the multiuser environment, such as near-far situations and high bandwidth efficiency utilization. In the presence of strong interference the proposed detector achieves the performance of the optimum detector. When interfering signals are weak, it performs slightly better than its non-adaptive counterpart (one that requires prior estimation of the received signal energies), but without requiring knowledge of the received signal energies. This paper deals with the synchronous case, which is not aplicable to base station processing. The asynchronous case, as an extension to the synchronous case approach, is currently being pursued by the authors.

V. APPENDIX

By definition:

$$\begin{aligned}
Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} &= Pr\{\mathrm{sgn}(\sqrt{a_1}b_1+\xi_1),\ldots,\mathrm{sgn}(\sqrt{a_{k-1}}b_{k-1}+\xi_{k-1}),\\
&\quad \mathrm{sgn}(\sqrt{a_{k+1}}b_{k+1}+\xi_{k+1}),\ldots,\mathrm{sgn}(\sqrt{a_K}b_K+\xi_K)\}\\
&= Pr\{\hat{b}_1\xi_1 > -\sqrt{a_1}b_1\hat{b}_1,\ldots,\hat{b}_{k-1}\xi_{k-1} > -\sqrt{a_{k-1}}b_{k-1}\hat{b}_{k-1},\\
&\quad \hat{b}_{k+1}\xi_{k+1} > -\sqrt{a_{k+1}}b_{k+1}\hat{b}_{k+1},\ldots,\hat{b}_K\xi_K > -\sqrt{a_K}b_K\hat{b}_K\}\\
&= Pr\{\hat{b}_1\xi_1 < \sqrt{a_1}b_1\hat{b}_1,\ldots,\hat{b}_{k-1}\xi_{k-1} < \sqrt{a_{k-1}}b_{k-1}\hat{b}_{k-1},\\
&\quad \hat{b}_{k+1}\xi_{k+1} < \sqrt{a_{k+1}}b_{k+1}\hat{b}_{k+1},\ldots,\hat{b}_K\xi_K < \sqrt{a_K}b_K\hat{b}_K\}\\
&= Pr\{-\hat{\boldsymbol{b}}_k|-\boldsymbol{b}_k\}
\end{aligned}$$

VI. ACKNOWLEDGEMENT

The authors are grateful to Zvi Drezner from California State University, Fullerton, for kindly providing the source code for DMV (Double precision MultiVariate integral) [7].

VII. REFERENCES

[1] S. Verdu, *"Minimum probability of error for asynchronous Gaussian multiple access channels,"* IEEE Trans. Inform. Theory, Vol. IT-32, No. 1, pp. 85-96, Jan. 1986.

[2] R. Lupas and S. Verdu, *"Linear multiuser detectors for synchronous code division multiple access channels,"* IEEE Trans. Inform. Theory, vol. IT-35, No. 1, pp. 123-136, Jan. 1989.

[3] R. Lupas and S. Verdu, *"Near far resistance of multiuser detectors in asynchronous channels,"* IEEE Trans Commun., vol. COM-38, No. 4, pp. 496-508, Apr. 1990.

[4] A. Duel-Hallen, *"Decorrelating decision-feedback multiuser detector for synchronous code division multiple access channel",* IEEE Trans. Commun., vol. COM-41, No. 2, pp. 285-290, Feb. 1993.

[5] M. K. Varanasi and B. Aazhang, *"Multistage detector in asynchronous code-division multiple access communications,"* IEEE Trans. Commun. vol. 38, No. 4, pp. 509-519, Apr. 1990.

[6] M. K. Varanasi and B. Aazhang, *"Near-optimum detector in synchronous code-division multiple-access system,"* IEEE Trans. Commun. vol. 39, No. 5, pp. 725-736, May 1991.

[7] Z. Drezner, *"Computation of the multivariate normal integral,"* ACM Trans. on Mathematical Software, Vol. 18, No. 4, pp. 470-480, Dec. 1992.

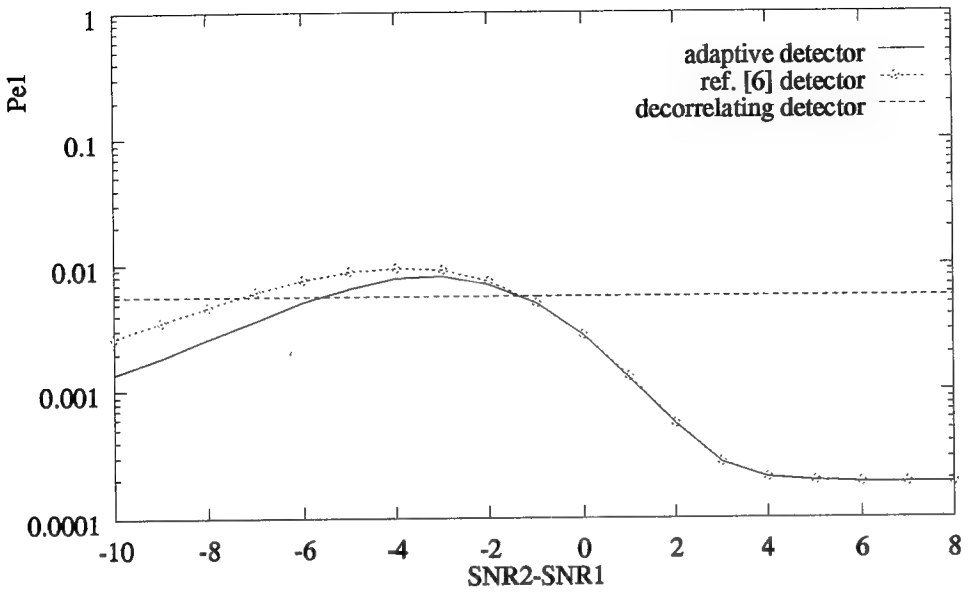


Fig. 2 Error probability of user 1 for K=2, ρ =0.7 and SNR1=8 dB

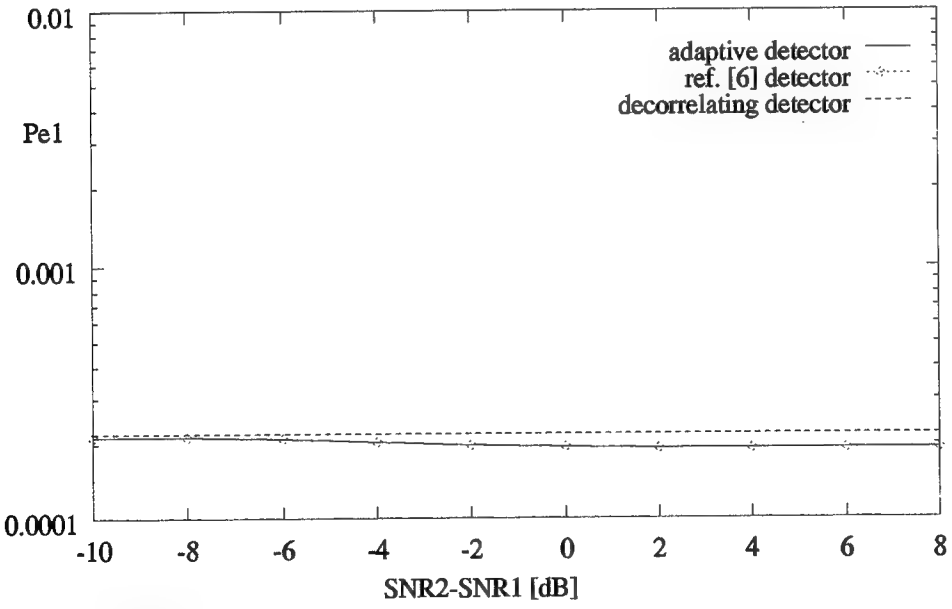


Fig. 3 Error probability of user 1 for K=2 and SNR1=8 dB

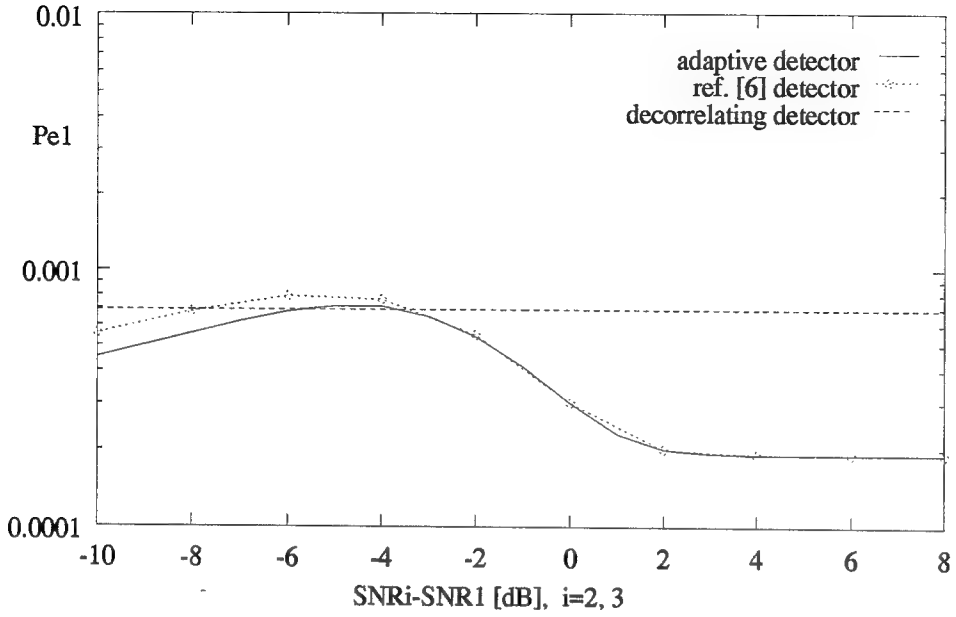


Fig. 4 Error probability of user 1 for K=3 and SNR1=8 dB

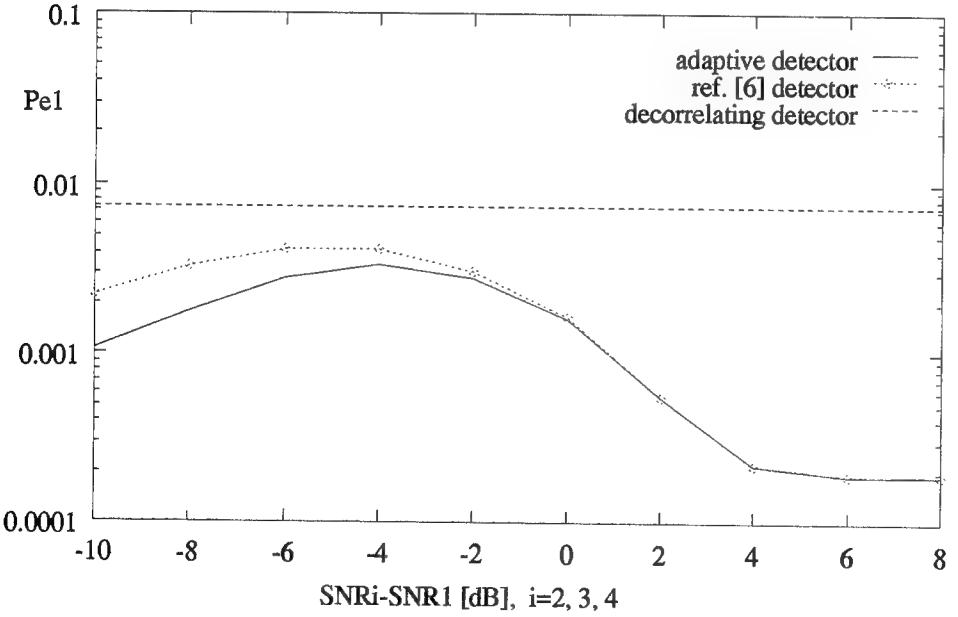


Fig. 5 Error probability of user 1 for K=4 and SNR1=8 dB

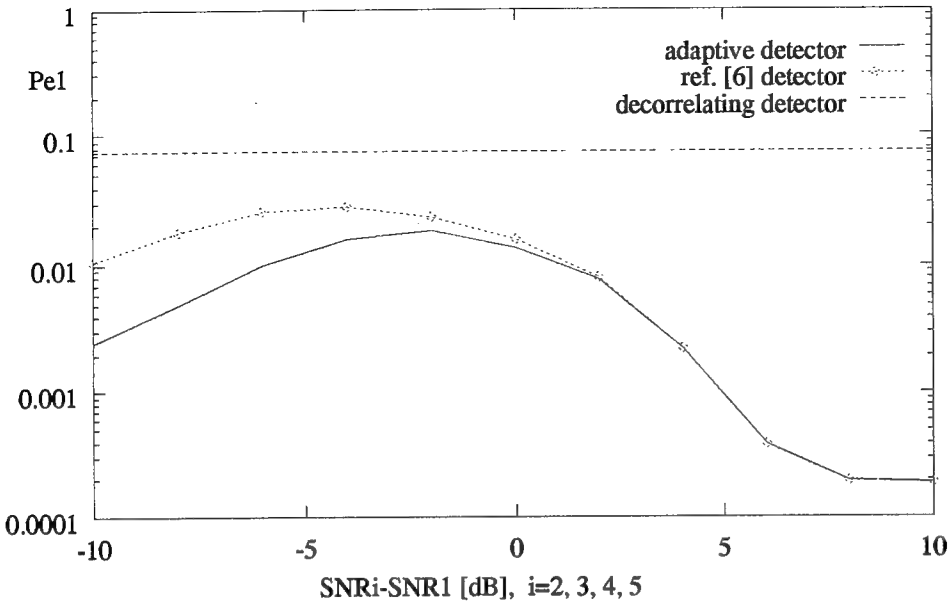


Fig. 6 Error probability of user 1 for K=5 and SNR1=8 dB

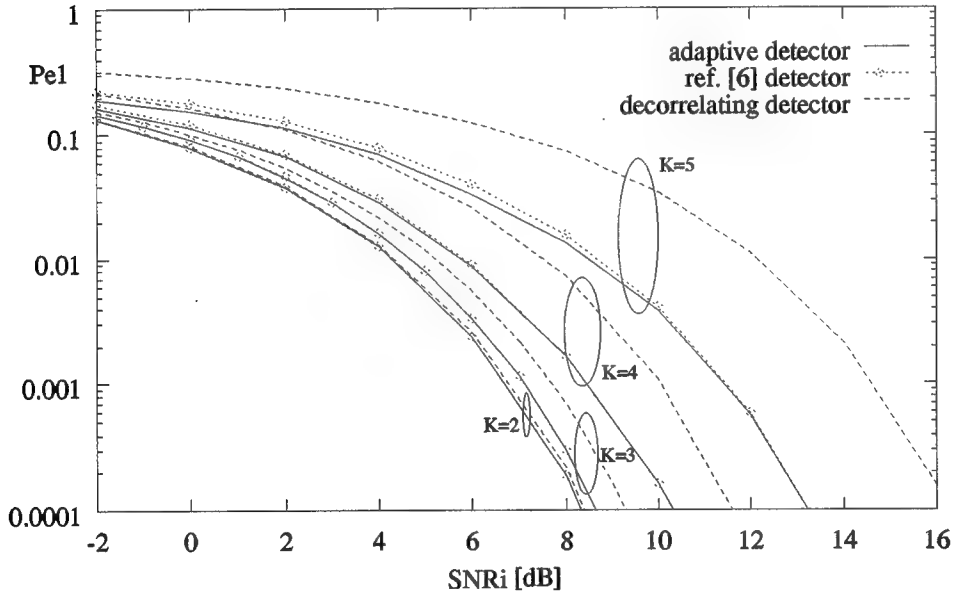


Fig. 7 Error probability of user 1 for K=2 to K=5

APPENDIX B: PART II

SYNCHRONOUS MULTIUSER CDMA DETECTOR WITH SOFT DECISON ADAPTIVE CANCELER

by
David W. Chen, Zoran Siveski and Yeheskel Bar-Ness

ABSTRACT

A two-stage multiuser detector for the synchronous code division multiple access (CDMA), additive Gaussian noise channel is considered. It employs a combination of a decorrelator and a nonlinear multiuser interference canceler that utilizes soft tentative decisions. The weights of the canceler are adaptively adjusted, thus rendering the knowledge of the received signal energies unnecessary. The steady state error performance of the detector is obtained and found to be superior to the performance of the same detector that utilizes hard tentative decisions.

I. INTRODUCTION

A synchronous CDMA receiver that employs a combination of a decorrelator and a multiuser interference canceler and whose weights are adaptively controlled was presented in [1]. The scheme did not require a prior estimate of received signal energies, as was the case with its non-adaptive, fixed weight counterpart described in [2]. The tentative data bit estimates used in the canceler were obtained as hard decisions on decorrelator outputs. In the presence of strong multiuser interference, in the steady state, the adaptive scheme approached the performance of an optimum receiver, while in the presence of weak interference it performed better than [2]. In this paper, soft tentative decisions, in the manner described in [3,4] are used instead, and the steady state error performance of the detector is evaluated.

II. DETECTOR: MODEL AND ANALYSIS

The detector considered here is shown in Fig. 1. The received signal $r(t)$ is expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i) \sqrt{a_k} s_k(t - iT) + n(t),$$

where $b_k(i) \in \{-1, +1\}$ is the k-th user's data bit in the i-th time interval and $n(t)$ is the additive, zero mean, white Gaussian noise with the two-sided power spectral density $N_0/2$. The received energy of the k-th user signal, unknown to the receiver, is denoted as a_k. The signature sequence $s_k(t)$, of the same duration T as a data bit, is known to the receiver. The sampled outputs of the bank of matched filters in the i-th bit interval can be expressed as:

$$\boldsymbol{x} = \boldsymbol{\mathcal{P}} \boldsymbol{A} \boldsymbol{b} + \boldsymbol{n},$$

where: $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]^T$, $\boldsymbol{b} = [b_1, b_2, \ldots, b_K]^T$,
$\boldsymbol{n} = [n_1, n_2, \ldots, n_K]^T$ and $\boldsymbol{A} = diag[\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_K}]$.

The (k, j)th element of the symmetric crosscorrelation matrix $\boldsymbol{\mathcal{P}}$ is defined as:

$$\rho_{kj} = \int_0^T s_k(t) s_j(t) dt,$$

with $\rho_{kk} = 1$. The outputs of the decorrelator are given as:

$$\boldsymbol{z} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\xi},$$

where $\boldsymbol{z} = [z_1, z_2, \ldots, z_K]^T$ and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_K]^T$.

The output for the k-th user can be expressed as:

$$y_k = x_k - \boldsymbol{w}_k{}^T \boldsymbol{h}_k = x_k - \sum_{\substack{l=1 \\ l \neq k}}^{K} w_{lk} h_{lk}, \tag{1}$$

where $\boldsymbol{w}_k$ is the k-th column vector of a $K \times K$ weight matrix $\boldsymbol{W}$, with the element $w_{kk} = 0$ deleted, and $\boldsymbol{h}_k$ is the k-th column vector of a $K \times K$ matrix $\boldsymbol{H}$, with the element $h_{kk} = 0$ deleted. The (l, k)th element of the matrix $\boldsymbol{H}$ represents the output of the soft limiter to the input z_l, and, for $l, k = 1, \ldots, K \quad l \neq k$, is defined as:

$$h_{lk}(z_l) = \begin{cases} z_l/\lambda_{lk} & |z_l| < \lambda_{lk} \\ sgn(z_l) & \text{otherwise.} \end{cases} \tag{2}$$

The limiter's parameter λ_{lk} is determined heuristically from the observed values of the decorrelator outputs as:

$$\lambda_{lk} = \frac{\rho_{lk}[E\{|z_k|\}]^2}{E\{|z_l|\}}, \tag{3}$$

where the above expectation is evaluated as:

$$E\{|z_k|\} = \sqrt{a_k}\left(1 - 2Q\left(\frac{\sqrt{a_k}}{\sigma_{\xi_k}}\right)\right) + \frac{2\sigma_{\xi_k}}{\sqrt{2\pi}} e^{-\frac{a_k}{2\sigma_{\xi_k}^2}},$$

where:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt,$$

and $\sigma_{\xi_k}^2$ is the variance of ξ_k.

For controlling the weights, the steepest descent algorithm, which simultaneously minimizes the output signal powers $E\{y_k^2\}$, is used. That is, for the k-th output, the optimum weights are obtained by the iterative search

$$\begin{aligned} \boldsymbol{w}_k(i+1) &= \boldsymbol{w}_k(i) - \frac{\mu}{2} \frac{\partial}{\partial \boldsymbol{w}_k(i)} E\{y_k^2(i)\} \\ &= (\boldsymbol{I} - \mu E\{\boldsymbol{h}_k \boldsymbol{h}_k^T\}) \boldsymbol{w}_k(i) + \mu E\{x_k \boldsymbol{h}_k\}. \end{aligned}$$

(In practice, the expectation operation is replaced by time averaging).
The steady state values can be evaluated from:

$$\frac{\partial}{\partial \boldsymbol{w}_k} E\{y_k^2\} = 0 = -E\{x_k \boldsymbol{h}_k\} + E\{\boldsymbol{h}_k \boldsymbol{h}_k^T\} \boldsymbol{w}_k. \tag{4}$$

It is easily shown that:

$$\begin{aligned} E\{x_k \boldsymbol{h}_k\} &= E\left\{\left[\sqrt{a_k} b_k + \boldsymbol{\rho}_k^T \boldsymbol{A}_k \boldsymbol{b}_k + n_k\right] \boldsymbol{h}_k\right\} \\ &= \left[\boldsymbol{A}_k E\{\boldsymbol{b}_k \boldsymbol{h}_k^T\}\right]^T \boldsymbol{\rho}_k, \end{aligned}$$

where $\boldsymbol{A}_k$ is a diagonal $(K-1) \times (K-1)$ submatrix of $\boldsymbol{A}$ with its k-th diagonal entry removed, $\boldsymbol{\rho}_k$ is a $(K-1) \times 1$ k-th column vector obtained from $\boldsymbol{\mathcal{P}}$ by deleting the element ρ_{kk}, and $\boldsymbol{b}_k$ is a $(K-1) \times 1$ vector obtained from $\boldsymbol{b}$ by deleting the element b_k. Clearly $E\{\boldsymbol{b}_k \boldsymbol{h}_k^T\}$ is diagonal; therefore, the system of $K-1$ linear equations (4) gives the steady state values of the weights affecting the k-th output as:

$$\boldsymbol{w}_k = \left[E\{\boldsymbol{h}_k \boldsymbol{h}_k^T\}\right]^{-1} \boldsymbol{A}_k E\{\boldsymbol{b}_k \boldsymbol{h}_k^T\} \boldsymbol{\rho}_k. \tag{5}$$

The expectations in the above expression are given in the Appendix.

Defining the k-th user final decision output as $\hat{b}_k = sgn(y_k)$, its probability of error is evaluated as follows:

$$\begin{aligned} P_{e_k} &= E_{b_k, \boldsymbol{b}_k, \boldsymbol{h}_k} Pr\{\hat{b}_k \neq b_k | b_k, \boldsymbol{b}_k, \boldsymbol{h}_k\} \\ &= \frac{1}{2^{K-1}} \sum_{\boldsymbol{b}_k} E_{\boldsymbol{h}_k} Pr\{-\sqrt{a_k} + \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k \\ &\quad -\boldsymbol{w}_k^T \boldsymbol{h}_k + n_k > 0 | \boldsymbol{h}_k\}. \end{aligned}$$

Introducing the transformed Gaussian random variable ψ_k,

$$\psi_k = n_k - \sum_{\substack{i=1 \\ i \neq k}}^{K} c_i \frac{w_{ik}}{\lambda_{ik}} \xi_i,$$

where

$$c_i = \begin{cases} 1 & |z_i| < \lambda_{ik} \\ 0 & \text{otherwise} \end{cases} \quad i = 1, 2, \ldots, K, \; i \neq k,$$

the error probability becomes:

$$\begin{aligned} &P_{e_k} = \frac{1}{2^{K-1}} \cdot \\ &\sum_{\boldsymbol{b}_k} E_{\boldsymbol{h}_k} Pr\{\psi_k > \sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \boldsymbol{g}_k | \boldsymbol{h}_k\}, \end{aligned}$$

where

$$\boldsymbol{g}_k = [g_{1k}, g_{2k}, \ldots, g_{k-1,k}, g_{k+1,k}, \ldots, g_K]^T \text{ with}$$

$$g_{ik} = \begin{cases} \sqrt{a_i}b_i/\lambda_{ik} & |z_i| < \lambda_{ik} \\ sgn(z_i) & \text{otherwise} \end{cases} \quad i = 1, 2, \ldots, K,\ i \neq k.$$

Defining the vector $\boldsymbol{\zeta}$,

$$\boldsymbol{\zeta} = [\xi_1, \xi_2, \ldots, \xi_{k-1}, \psi_k, \xi_{k+1}, \ldots, \xi_K]^T,$$

the final expression for the error probability is obtained as:

$$P_{e_k} = 2^{-(K-1)} \sum_{\boldsymbol{b}_k} \sum_{n=1}^{2^{K-1}} \int_{\boldsymbol{D}_n} \boldsymbol{f}_{\boldsymbol{\zeta}} d\boldsymbol{\zeta}, \tag{6}$$

where $\boldsymbol{f}_{\boldsymbol{\zeta}}$ is a K-variate Gaussian density function, $\boldsymbol{D}_n$ is a hyper cube defined by: $\sqrt{a_k} - \boldsymbol{b}_k^T \boldsymbol{A}_k \boldsymbol{\rho}_k + \boldsymbol{w}_k^T \boldsymbol{g}_k < \zeta_k$ and for $i = 1, 2, \ldots, K$, $i \neq k$,

$$\begin{array}{ll} -\lambda_{ik} - \sqrt{a_i}b_i < \zeta_i < \lambda_{ik} - \sqrt{a_i}b_i & |z_i| < \lambda_{ik} \\ \zeta_i > \lambda_{ik} - \sqrt{a_i}b_i,\ \zeta_i < -\lambda_{ik} - \sqrt{a_i}b_i & \text{otherwise.} \end{array}$$

III. NUMERICAL EXAMPLES AND DISCUSSION

The steady state error probabilities are presented for the two and three-user cases. The signal to noise ratio for user k (SNR_k) is defined as a_k/N_0. In Fig. 2, SNR_1 is set to 8 dB while SNR_2/SNR_1 varies from -10 to 8 dB. The value of the crosscorrelation coefficient $\rho_{12} = 0.7$, which represents a high bandwidth efficiency case. Compared to the error performance of the detector from [1], which utilizes hard decisions, and the performance of the decorrelator, a significant improvement has been obtained.

In the next two examples, Gold codes of length seven (e.g., [2] and [1]) were used for signature waveforms. For the two-user case, in a low bandwidth efficiency case with $\rho_{12} = -1/7$ (Fig. 3) as expected, a performance of the decorrelator is very close to the single user bound. Therefore, a negligible improvement is obtained with either the hard or soft tentative decisions. The three-user case with $\rho_{13} = 3/7$ and $\rho_{23} = -1/7$ is depicted in Fig. 4. Here, the two-stage detector with a soft limiter clearly shows the best performance. For a different set of crosscorrelation coefficient values ($\rho_{12} = 0.5$, $\rho_{13} = 0.5$, and $\rho_{23} = 0.2$) in Fig. 5, the difference in the performance between the decorrelator and the two-stage detector is larger, with the soft limiter again being a better one. Finally, Fig. 6 depicts the error probabilities in the case when all the users maintain the same $SNR's$ and the Gold codes are used.

IV. APPENDIX

Let f_{z_i} be the density function of z_i, and $\boldsymbol{f}_{z_i z_j}$ the joint density function of z_i and z_j. Then:

$$E\{h_{ik}^2\} = \frac{1}{2} \sum_{b_i} \left[\int_{\boldsymbol{L}_1} \frac{z_i^2}{\lambda_{ik}^2} f_{z_i} dz_i + \int_{\boldsymbol{L}_2} f_{z_i} dz_i \right]$$

$$
\begin{aligned}
E\{h_{ik}h_{jk}\} &= \frac{1}{4}\sum_{b_i,b_j}\Big[\int\int_{\boldsymbol{Z}_1}\frac{z_i z_j}{\lambda_{ik}\lambda_{jk}}\boldsymbol{f}_{z_i z_j}dz_i dz_j \\
&+ \int\int_{\boldsymbol{Z}_2}\frac{z_i}{\lambda_{ik}}sgn(z_j)\boldsymbol{f}_{z_i z_j}dz_i dz_j \\
&+ \int\int_{\boldsymbol{Z}_3}\frac{z_j}{\lambda_{jk}}sgn(z_i)\boldsymbol{f}_{z_i z_j}dz_i dz_j \\
&+ \int\int_{\boldsymbol{Z}_4}sgn(z_i)sgn(z_j)\boldsymbol{f}_{z_i z_j}dz_i dz_j\Big]
\end{aligned}
$$

$$
\begin{aligned}
E\{b_i h_{ik}\} &= \frac{1}{2}\sum_{b_i}\Big[\int_{\boldsymbol{L}_1} b_i\frac{z_i}{\lambda_{ik}}f_{z_i}dz_i \\
&+ \int_{\boldsymbol{L}_2} b_i sgn(z_i)f_{z_i}dz_i\Big],
\end{aligned}
$$

where $\boldsymbol{L}_i$, $i = 1, 2$ correspond to the appropriate intervals of z_i, and $\boldsymbol{Z}_i$, $i = 1, \ldots, 4$ correspond to the appropriate rectangular regions in the (z_i, z_j) plane.

References

[1] Z. Siveski, Y. Bar-Ness and D. W. Chen, "Adaptive multiuser detector for synchronous code division multiple access applications," 1994 *International Zurich Seminar on Digital Communications*, ETH Zurich, Switzerland, Mar. 1994.

[2] M. K. Varanasi and B. Aazhang, "Near optimum detector in synchronous code division multiple access system," *IEEE Trans. Commun.*, vol. 39, No. 5, pp. 725-736, May 1991.

[3] X. Zhang and D. Brady, "Soft-decision multistage detection for asynchronous AWGN channels," *31st Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sept. 1993.

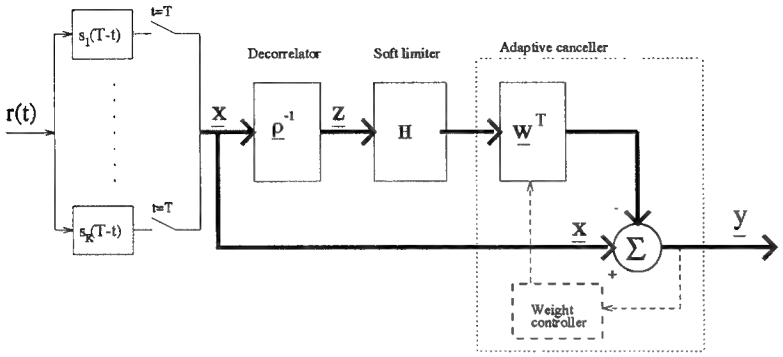


Figure 1: Figure 1: Proposed detector scheme

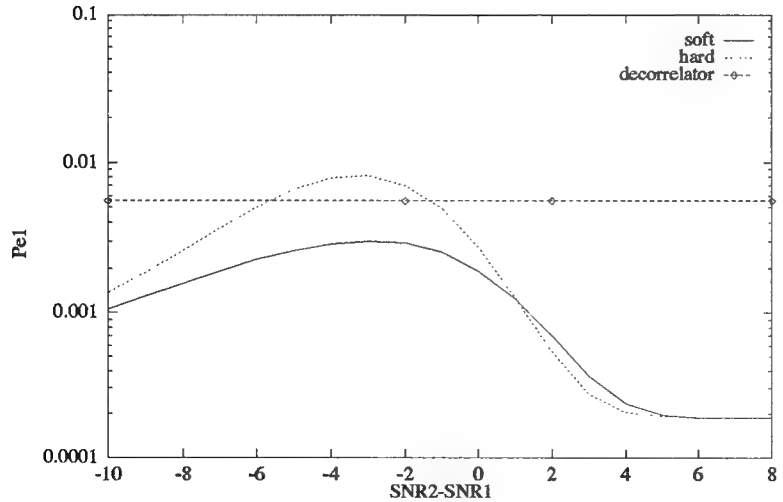


Figure 2: Error probability of user 1 for $K = 2$, and $SNR_1 = 8$ dB and $\rho_{12} = 0.7$.

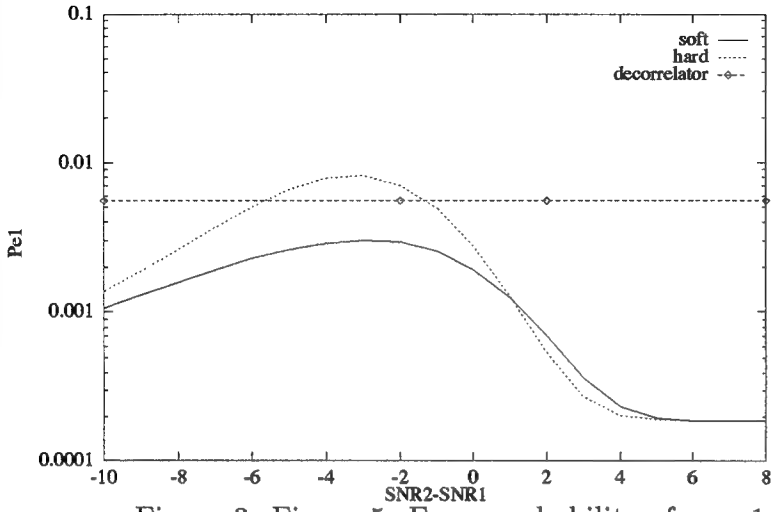


Figure 3: Figure 5: Error probability of user 1 for $K = 3$

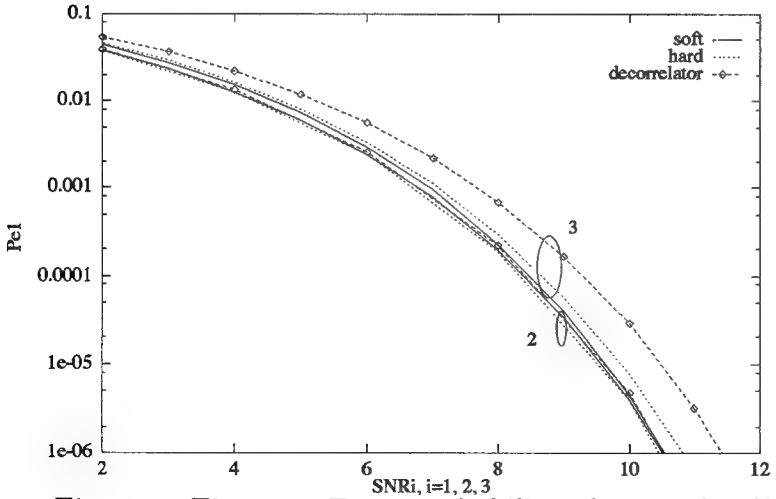


Figure 4: Figure 6: Error probability of user 1 for $K = 2$ and $K = 3$

APPENDIX B: PART III

BOOTSTRAPPED DECORRELATING ALGORITHM FOR ADAPTIVE INTERFERENCE CANCELATION IN SYNCHRONOUS CDMA COMMUNICATIONS SYSTEMS

by
Yeheskel Bar-Ness, Zoran Siveski and David W. Chen

ABSTRACT

A code division multiple access detector that employs a combination of a correlation detector and a multiuser adaptive interference canceler was previously proposed by the authors. The weights of the canceler are adaptively controlled by the steepest descent algorithm using the minimum power criterion of the canceler outputs. It was shown that such a canceler substantially outperforms the decorrelating detector and has almost the same error probability as a canceler that uses reliable estimates of the received signal energies. In this paper, a different weight control criterion based on minimizing the correlation between the signals at the outputs of the canceler is used, and its performance is compared to that obtained with the minimum power criterion. It is shown that both cancelers have almost the same performance, particularly when SNR is modestly high. Although it is not shown in this paper, we believe that minimizing correlation rather than power outputs is a more suitable criterion when dealing with highly dispersive channels.

I. INTRODUCTION

A conventional single user detector implemented in the Code Division Multiple Access (CDMA) system consists of a bank of filters, each one matched to the signature sequence of the particular user. The sampled output of each matched filter, besides the desired signal, contains the residual interference from all other users. In particular, for mobile and wireless personal communication the choice of CDMA is attractive because of its potential capacity increase and other technical factors, such as anti-multipath fading capability. However, for satisfactory performance, one must consider the effect of the "near-far" problem. So far, practical systems have used power control to combat this problem. Commercial digital cellular systems based on CDMA, and which use stringent power control are described in [1]. They offer capacity increases over TDMA. A different approach for combating the near-far problem was suggested by Verdu [2], where, an optimum receiver in the multiuser interference environment was proposed and shown to eliminate the near-far problem and provide much improved performance. The improvement comes at the expense of high computational complexity. A class of suboptimum receivers that uses decorrelating detectors and which is based on the linear transformation of the sampled matched filters' outputs was considered in [3] and [4]. The decorrelating decision-feedback detector presented in [5] utilizes the

difference in received users' energies where the decisions of the stronger users are used to eliminate interference on weaker users.
Another approach for suboptimum multiuser detectors with low complexity was proposed in [6] and [7]. The performance of some of these suboptimum schemes is close to the performance of the optimum detector. Particularly when the power of the interferes increase, they become indistinguishable. However, these schemes have to perform an estimation of the received signal energies, knowledge of which is required for the detector's proper operation.
Still another approach for improving the performance of CDMA systems in multiuser environments is implementing adaptive algorithms for signal separation [8, 9].
In contrast to conventional interference cancelers, wherein the interference to desired signal ratio is high, at the output of the matched filter of a Direct Sequence Code Division Multiple Access (DS/CDMA), the desired signal is higher and the designer's aim is to further reduce the interference. Unless special arrangements are made, conventional cancelers are useless for removing interference in CDMA systems. Such an arrangement includes, for example, the use of a reference in the form of an estimate of the desired signal, which can be obtained by a training sequence, etc. Clearly, such an approach is difficult, particularly in vehicular communication. Variable desired signal-to-interference ratios resulting from the "near-far" problem make the difficulty even worse. Instead, we used in [9] and in this paper, the tentative decision on the data following the decorrelation transformation (see Figure 1).
When dealing with an adaptive scheme, the questions are what kind of performance index and which algorithm to use. While the minimum power criterion was implemented in [9], here we propose to use a special form of decorrelating algorithm frequently used in neural networks and other applications of signal separation [10,11].
In Section 2, the system model of the adaptive canceler will be described and the decorrelating algorithm will be presented. An analytical expression for the equations from which the steady state mean value of the controlled weight will be given and simplified using some approximation particularly valid for reasonably high SNR's. The output error probability is calculated in Section 3 (using the aforementioned approximation), then exact computation of this probability for the case of two users is given in Section 4. Section 5 contains computational and simulation results for different signal's scenarios, after which some conclusions are drawn.

II. THE ADAPTIVE CANCELER: MODEL AND ANALYSIS

A synchronous CDMA receiver considered here consists of a bank of filters (matched to the users' signature sequences) front-end, followed by samplers and the decision system, as depicted in Fig. 1.
The received signal $r(t)$ is expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i) \sqrt{a_k} s_k(t - iT) + n(t) \tag{1}$$

where $b_k(i) \in \{-1, +1\}$ is the k-th user's data bit in the i-th time interval and $n(t)$ is the additive, zero mean, white Gaussian noise with the two-sided power spectral density $N_0/2$.

The received energy of the k-th user signal, unknown to the receiver, is denoted as a_k. The signature sequence $s_k(t)$, of the same duration T as a data bit, is known to the receiver. The sampled outputs of the bank of matched filters in the i-th bit interval can be expressed as:

$$\boldsymbol{x}(i) = \boldsymbol{\mathcal{P}}\boldsymbol{A}\boldsymbol{b}(i) + \boldsymbol{n}(i) \tag{2}$$

For the sake of convenience, the index i will be omitted elsewhere in the text. In (2), $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]^T$, $\boldsymbol{b} = [b_1, b_2, \ldots, b_K]$, $\boldsymbol{A} = \text{diag}[\sqrt{a_1}, \sqrt{a_2}, \ldots, \sqrt{a_K}]$ is a diagonal matrix and $\boldsymbol{n} = [n_1, n_2, \ldots, n_K]^T$. The (k, j)th element of the symmetric cross-correlation matrix $\boldsymbol{\mathcal{P}}$ is defined as:

$$\rho_{kj} = \int_0^T s_k(t)s_j(t)dt \qquad k, j \in (1, 2, \ldots, K)$$

with $\rho_{kk} = 1$. The covariance matrix of a zero mean Gaussian noise vector $\boldsymbol{n}$ is,

$$E\{\boldsymbol{n}\boldsymbol{n}^T\} = \frac{N_0}{2}\boldsymbol{\mathcal{P}}$$

The outputs of the decorrelator are given as:

$$\boldsymbol{z} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\Gamma}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\xi} \tag{3}$$

where $\boldsymbol{\Gamma}$ is the inverse of $\boldsymbol{\mathcal{P}}$, and $\boldsymbol{\xi} = \boldsymbol{\Gamma}\boldsymbol{n}$. Let $\hat{\boldsymbol{b}} = \left[\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_K\right]^T$ be the decision output of the decorrelator defined as:

$$\hat{\boldsymbol{b}} = \text{sgn}(\boldsymbol{z}) \tag{4}$$

Then the canceler's output is given by:

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{W}^T\hat{\boldsymbol{b}} \tag{5}$$

where

$$\boldsymbol{W} = \begin{bmatrix} 0 & w_{12} & \ldots & w_{1K} \\ w_{21} & 0 & \ldots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \ldots & 0 \end{bmatrix}$$

The output for the k-th user can be expressed as:

$$y_k = x_k - \boldsymbol{w}_k{}^T\hat{\boldsymbol{b}}_k \tag{6}$$

where $\boldsymbol{w}_k$ is the k-th column vector of $\boldsymbol{W}$ with the element w_{kk} deleted, and $\hat{\boldsymbol{b}}_k$ is the vector obtained from $\hat{\boldsymbol{b}}$ by deleting the element $\hat{b}_k$. The multidimensional canceler structure is the same as the one termed in an array processing sidelobe canceler, except for the fact that for auxiliary signals we are using the decision outputs of the decorrelator.

For controlling the weights, we use the steepest descent algorithm, which simultaneously reduces the absolute value of the correlation between the data output $\hat{\hat{\boldsymbol{b}}}_m$ of channel m and all other outputs of the canceler (before decision). That is the weight w_{ml} is controlled by the recursion

$$w_{ml} \leftarrow w_{ml} - \mu E\{y_l \mathrm{sgn}(y_m)\} \quad \text{for } 1 \le l, m \le K \quad l \ne m \tag{7}$$

To see the logic behind this recursion for controlling the weight we refer to Fig. 2, which is the special case of three users. Recursion (7) reaches steady state in the mean when $E\{y_l \mathrm{sgn}(y_m)\} = 0$. Notice that w_{21}, for example, is used to cancel residue of data b_2 at the output y_1. It will settle down only when that residue (being correlated with $\hat{\hat{\boldsymbol{b}}}$) is zero. But any reduction of b_2 at y_1 will improve $\hat{\hat{\boldsymbol{b}}}_1$ (smaller error) and, hence, will be more effective in reducing the residue of b_1 at y_3 and y_2 through the control weights w_{12} and w_{13}. Therefore, the process of residue cancelation is enhanced successively, which justifies the name "bootstrap." In vector form, by using the notation of Eq. 6, we write:

$$\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k - \mu E\{y_k \mathrm{sgn}(\boldsymbol{y}_k)\} \tag{8}$$

where again $\boldsymbol{y}_k$ is obtained from the column vector $\boldsymbol{y}$ by deleting the element y_k.
The steady state would be reached if

$$E\{y_k \mathrm{sgn}(\boldsymbol{y}_k)\} = 0, \quad \text{for } k = 1, 2, \ldots, K$$

This means that $K(K-1)$ nonlinear equations have to be evaluated. (An example of $K = 2$ for finding the steady state values of w_{21} and w_{12} will be given in the next section).
In the general case, it will be assumed that the signal-to-noise ratios (SNR's) are large enough such that the main contribution to the output error is the multiuser interference.
Therefore, $E\{y_k \mathrm{sgn}(\boldsymbol{y}_k)\} = E\{y_k \hat{\hat{\boldsymbol{b}}}_k\}$ can be approximated as:

$$\begin{aligned} E\{y_k \hat{\hat{\boldsymbol{b}}}_k\} &\approx E\{y_k \boldsymbol{b}_k\}(1 - 2Pr(\hat{\hat{\boldsymbol{b}}}_k \text{ in error})) \\ &= E\{y_k \boldsymbol{b}_k\}(1 - 2\boldsymbol{P}_{e_k}) \end{aligned} \tag{9}$$

Substituting (6) into (9) with x_k taken from (2), we get:

$$\begin{aligned} E\{y_k \hat{\hat{\boldsymbol{b}}}_k\} \approx\ & E\left\{\left(\sqrt{a_k} b_k + \boldsymbol{\rho}_k^T \boldsymbol{A}_k \boldsymbol{b}_k + n_k - \boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\right) \boldsymbol{b}_k\right\} \\ & \cdot(1 - 2\boldsymbol{P}_{e_k}\ ), \end{aligned}$$

and the algorithm will result in having:

$$\begin{aligned} & \sqrt{a_k} E\left\{b_k \boldsymbol{b}_k\right\} + E\left\{\left(\boldsymbol{\rho}_k^T \boldsymbol{A}_k \boldsymbol{b}_k\right) \boldsymbol{b}_k\right\} \\ & + E\left\{n_k \boldsymbol{b}_k\right\} - E\left\{\left(\boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\right) \boldsymbol{b}_k\right\} = 0 \end{aligned}$$

It is easy to show that $E\left\{b_k \boldsymbol{b}_k\right\} = 0$, $E\left\{\left(\boldsymbol{\rho}_k^T \boldsymbol{A}_k \boldsymbol{b}_k\right) \boldsymbol{b}_k\right\} = \boldsymbol{A}_k^T \boldsymbol{\rho}_k$, $E\left\{n_k \boldsymbol{b}_k\right\} = 0$, and $E\left\{\left(\boldsymbol{w}_k^T \hat{\boldsymbol{b}}_k\right) \boldsymbol{b}_k\right\} = \boldsymbol{B}_k \boldsymbol{w}_k$, where $\boldsymbol{B}_k = \mathrm{diag}\left[E\left\{b_j \hat{b}_j\right\}\right]$, $j = 1, 2, \ldots, K \quad j \ne k$. Therefore,

$$\boldsymbol{w}_k = \boldsymbol{B}_k^{-1} \boldsymbol{A}_k \boldsymbol{\rho}_k \tag{10}$$

where we used the fact that $\boldsymbol{A}_k$ is diagonal.

III. OUTPUT ERROR PROBABILITY

The output error probability for the k-th user is evaluated as follows:

$$\begin{aligned} P_{e_k} &= E_{b_k,\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k} Pr\{\hat{\hat{b}}_k \text{ in error}|b_k,\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k\} \\ &= \frac{1}{2}\sum_{\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k}\Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T\boldsymbol{A}_k\boldsymbol{\rho}_k + \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k\} \\ &+ Pr\{n_k < -\sqrt{a_k} - \boldsymbol{b}_k^T\boldsymbol{A}_k\boldsymbol{\rho}_k + \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k\}\Big]\cdot \\ &\quad Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\}Pr\{\boldsymbol{b}_k\} \end{aligned}$$

Since $Pr\{\boldsymbol{b}_k\} = 2^{-(K-1)}$ and n_k is a zero mean Gaussian random variable we can write:

$$\begin{aligned} P_{e_k} &= 2^{-K}\sum_{\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k}\Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T\boldsymbol{A}_k\boldsymbol{\rho}_k + \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k\} \\ &+ Pr\{n_k > \sqrt{a_k} + \boldsymbol{b}_k^T\boldsymbol{A}_k\boldsymbol{\rho}_k - \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k\}\Big]\, Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} \end{aligned} \tag{11}$$

Also

$$\begin{aligned} & Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} \\ = \; & Pr\{\mathrm{sgn}(\sqrt{a_1}b_1 + \xi_1),\ldots,\mathrm{sgn}(\sqrt{a_{k-1}}b_{k-1} + \xi_{k-1}), \\ & \mathrm{sgn}(\sqrt{a_{k+1}}b_{k+1} + \xi_{k+1}),\ldots,\mathrm{sgn}(\sqrt{a_K}b_K + \xi_K)\} \\ = \; & Pr\{\hat{b}_1\xi_1 > -\sqrt{a_1}b_1\hat{b}_1,\ldots,\hat{b}_{k-1}\xi_{k-1} > \\ & -\sqrt{a_{k-1}}b_{k-1}\hat{b}_{k-1},\hat{b}_{k+1}\xi_{k+1} > -\sqrt{a_{k+1}}b_{k+1}\hat{b}_{k+1}, \\ & \ldots,\hat{b}_K\xi_K > -\sqrt{a_K}b_K\hat{b}_K\} \\ = \; & Pr\{\hat{b}_1\xi_1 < \sqrt{a_1}b_1\hat{b}_1,\ldots,\hat{b}_{k-1}\xi_{k-1} < \sqrt{a_{k-1}}b_{k-1}\hat{b}_{k-1}, \\ & \hat{b}_{k+1}\xi_{k+1} < \sqrt{a_{k+1}}b_{k+1}\hat{b}_{k+1},\ldots,\hat{b}_K\xi_K < \sqrt{a_K}b_K\hat{b}_K\} \\ = \; & Pr\{-\hat{\boldsymbol{b}}_k|-\boldsymbol{b}_k\} \end{aligned} \tag{12}$$

Applying the result of (12) to the second term in (11) enables splitting the latter into two equal terms. Therefore

$$\begin{aligned} P_{e_k} &= \frac{1}{2^{K-1}}\sum_{\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k}\Big[Pr\{n_k > \sqrt{a_k} - \boldsymbol{b}_k^T\boldsymbol{A}_k\boldsymbol{\rho}_k + \boldsymbol{w}_k^T\hat{\boldsymbol{b}}_k\}\Big] \\ &\quad Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\} \end{aligned} \tag{13}$$

Substituting (10) in to (13) we finally obtain

$$P_{e_k} = \frac{1}{2^{K-1}}\cdot$$

$$\sum_{\boldsymbol{b}_k,\hat{\boldsymbol{b}}_k} Q\left(\frac{\sqrt{a_k} - \left(\boldsymbol{b}_k - \hat{\boldsymbol{b}}_k\boldsymbol{B}_k^{-1}\right)^T\boldsymbol{A}_k\boldsymbol{\rho}_k}{\sqrt{N_0/2}}\right) Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\}$$

where $Pr\{\hat{\boldsymbol{b}}_k|\boldsymbol{b}_k\}$ is the integral of the $(K-1)$ variate Gaussian density function.

The Two-user Case

For this case, one can solve for the optimal weight without the need for the approximation of (9), although the resulting two simultaneous equations in w_{21}, w_{12} are highly nonlinear and require numerical method for its solution. The calculated weight can then be substituted in the following equation to obtain the error probability:

$$P_{e1}(\epsilon) = \frac{1}{2}\left\{\left[Q\left(\frac{\sqrt{A_1}-\rho\sqrt{A_2}-w_{21}}{\sqrt{N_0/2}}\right)+ Q\left(\frac{\sqrt{A_1}+\rho\sqrt{A_2}+w_{21}}{\sqrt{N_0/2}}\right)\right] Q\left(\sqrt{\frac{(1-\rho^2)A_2}{N_0/2}}\right)+ \left[Q\left(\frac{\sqrt{A_1}+\rho\sqrt{A_2}-w_{21}}{\sqrt{N_0/2}}\right)+Q\left(\frac{\sqrt{A_1}-\rho\sqrt{A_2}+w_{21}}{\sqrt{N_0/2}}\right)\right] \cdot\left[1-Q\left(\sqrt{\frac{(1-\rho^2)A_2}{N_0/2}}\right)\right]\right\}$$

where,

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$$

IV. COMPUTATIONAL AND SIMULATION RESULTS

Two sets of examples are used to examine the performance of the proposed canceler. In the first set, we consider the two-user case with cross correlation coefficient $\rho_{12} = 0.7$. It can certainly represent a high bandwidth efficiency case. Here we find by computation the error probability as a function of the SNR's difference of the two users while the SNR of user 1 is kept constant at 8 dB and 12 dB. This is shown in Fig. 3 and Fig. 4, respectively. In the figures, we added the results of simulation, those obtained by using the approximation introduced in Eq.(9), and those resulting from the decorrelating detector of Verdu [3].
In the second set of examples, Gold sequences of length seven (frequently used in the literature, e.g., [4] and [6]) are chosen for signature waveforms. The rationale for such a choice is that Gold sequences are regularly used in an asynchronous CDMA environment and the study of its proposed synchronous counterpart may provide a useful indication of the performance of the former. The cross-correlation matrix $\boldsymbol{\mathcal{P}}$ in this case is:

$$\boldsymbol{\mathcal{P}} = \frac{1}{7}\begin{bmatrix} 7 & -1 & 3 & 3 & 3 \\ -1 & 7 & -1 & 3 & -1 \\ 3 & -1 & 7 & -1 & -1 \\ 3 & 3 & -1 & 7 & -1 \\ 3 & -1 & -1 & -1 & 7 \end{bmatrix}$$

In Fig. 5, we depict the result of error probability of user 1 having SNR=8dB as a function of the SNR of the other four users (taken to be the same). These results are obtained by simulations and computation with the approximation suggested in Eq.(9). For comparison, the results of the decorrelating detector are also included. Fig. 6 is the same except for user 1, where SNR=12 dB. Finally, Fig. 7 depicts the error probabilities of user 1, whose interferences are determined by the first column of $\mathcal{P}$ and having the same power as the other one to four interfering users. For comparison purposes we also show the results obtained by using the minimum power criterion and, as reference, those obtained when using the decorrelating detector. It was established that the case of equal user power represents approximately the worst case scenario for user 1. Also notice from $\mathcal{P}$ that other users suffer lower levels of interference by other users.

V. CONCLUSION

An adaptive interference canceler for the multiuser CDMA system is proposed in this paper. For the control of its adaptive weights, we suggested using a special form of the outputs' decorrelating algorithm, customarily used in neural networks and other applications of blind signal separation and equalization. It was shown that, particularly for a high level of interfering signals, the canceler outperforms the decorrelating detector, which is regularly suggested as a mean for combating the near-far problem without excessive hardware complexity. Although it is difficult to tackle the algorithm analytically, an approximation, suggested in this paper, made the task possible. Both the algorithm in this paper and the minimum output powers algorithm, previously proposed by the authors, depict similar error probability performance, satisfactory for candidacy in controlling adaptive interference cancelers for CDMA systems. Although we assume synchronous data signals in both cases, extension to the asynchronous case is also possible. Examination of the performance of these two cancelers for dispersive channels and, in particular, the difference in behavior of the two algorithms in such situations is the topic of current research.

References

[1] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, and C. Wheatley III, "On the capacity of cellular CDMA system," *IEEE Trans. on Vechicular Technology,* vol. VT-40, No. 2, pp. 303-311, May 1991.

[2] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple access channels," *IEEE Trans. Inform. Theory,* vol. IT-32, No. 1, pp. 85-96, Jan. 1986.

[3] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code division multiple access channels," *IEEE Trans. Inform. Theory,* vol. IT-35, No. 1, pp. 123-136, Jan. 1989.

[4] R. Lupas and S. Verdu, "Near far resistance of multiuser detectors for synchronous channels," *IEEE Trans. Comm.,*, vol. COM-38, No. 4, pp. 496-508, April 1990.

[5] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code division multiple access channel," *IEEE Trans. Comm.,* vol. COM-41, pp. 285-290, Feb. 1993.

[6] M.K. Varanasi and B. Aazhang, "Multistage detector in asynchronous code division multiple access communications," *IEEE Trans. Comm.,* vol. COM-38, No. 4, pp. 509-519, April 1990.

[7] M.K. Varanasi and B. Aazhang, "Near-optimum detector in synchronous code division multiple access system," *IEEE Trans. Comm.,* vol. COM-39, No. 5, pp. 725 -736, May 1991.

[8] Z. Siveski and Y. Bar-Ness, "Adaptive two-stage detection scheme in synchronous two-user CDMA systems," *MILCOM '93,* pp. 774-778, October 1993, Boston, MA.

[9] Z. Siveski, Y. Bar-Ness and D.W. Chen, "Error performance of synchronous multiuser code division multiple access detector with multidimensional adaptive canceler," accepted for publication in *European Transactions on Telecommunications and Related Technologies.*

[10] A. Dinc and Y. Bar-Ness, "Convergence and performance comparison of three different structures of bootstrap adaptive algorithm for multisignal co-channel separation," *MILCOM '92,* pp. 913-918, October 1992, San Diego, CA.

[11] R. Kamel and Y. Bar-Ness, "Blind decision feedback equalization using decorrelation criterion," *Communication Theory Mini Conference, Globecom '93,* December 1993, Houston, TX.

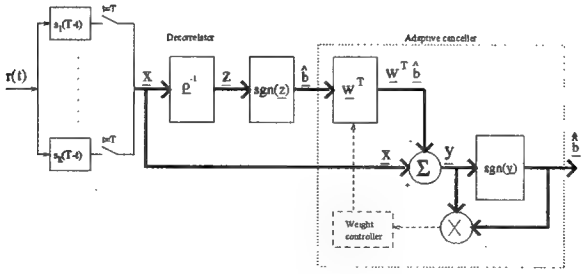


Fig. 1 Proposed multiuser receiver scheme

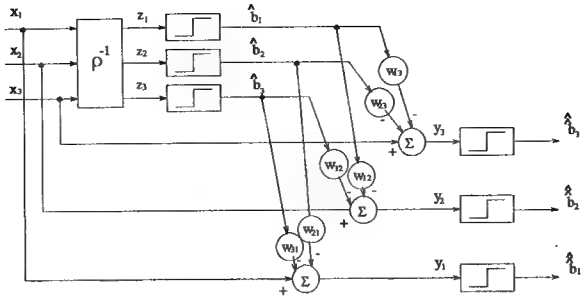


Fig. 2 Three-user canceler

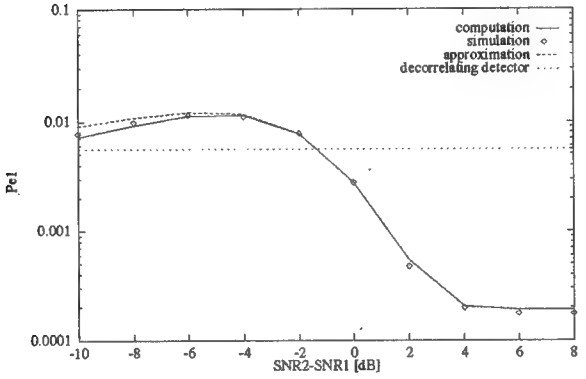


Fig.3 Probability of error for user 1 (K=2, $SNR_1 = 8$ dB, $\rho_{12} = 0.7$)

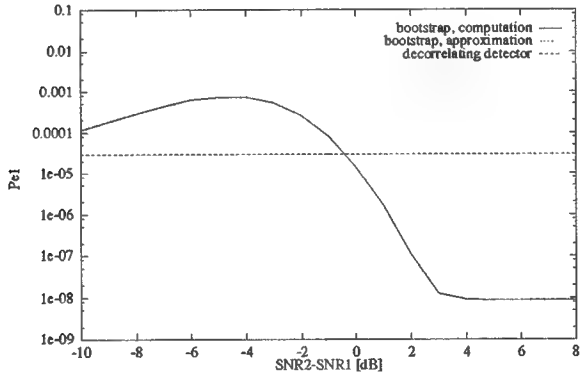


Fig.4 Probability of error for user 1 (K=2, $SNR_1 = 12$ dB, $\rho_{12} = 0.7$)

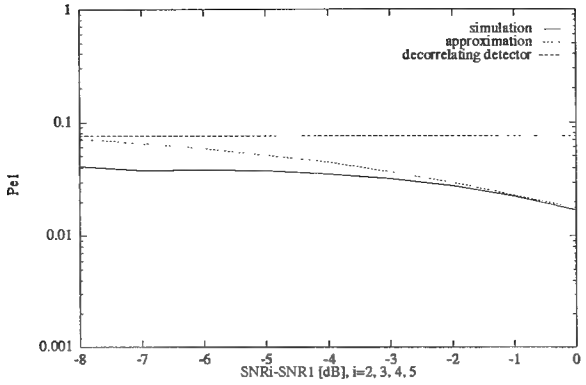


Fig.5 Probability of error for user 1 (K=5, $SNR_1 = 8$ dB, Gold codes)

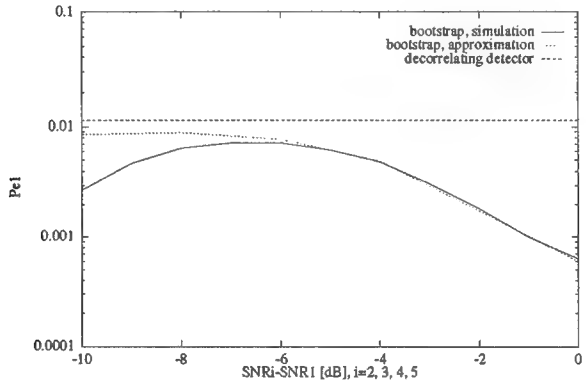


Fig.6 Probability of error for user 1 (K=5, $SNR_1 = 12$ dB, Gold codes)

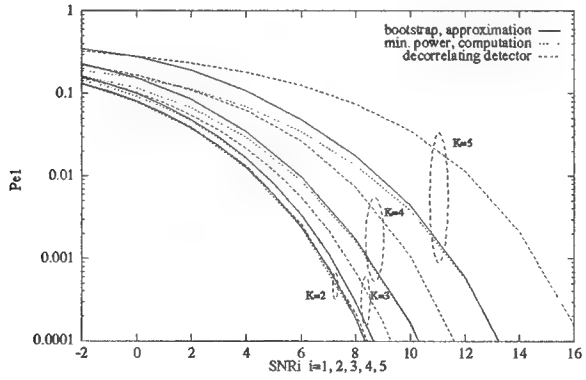


Fig.7 Probability of error for user 1 (K=5, Gold codes)

APPENDIX B: PART IV

ADAPTIVE MULTIUSER CDMA DETECTOR FOR ASYNCHRONOUS AWGN CHANNELS

by
Zoran Siveski, Lizhi Zhong and Yeheskel Bar-Ness

ABSTRACT

A multiuser detector in an additive white Gaussian noise, code division multiple access channel is proposed and analyzed. The detector utilizes the decorrelated tentative decisions obtained in the first stage to estimate and subtract multiple access interference in the second stage. The novel feature of the detector is the adaptive manner in which the multiple access interference estimates are formed, which renders prior estimation of the received signal amplitudes uneccessary. The detector is near-far resistant, and is shown to provide substantial performance improvement over the conventional and decorrelating detector, particularly in the presence of strong interfering signals.

I. PRELIMINARIES

In a multiuser environment, K users share the same channel with the unit energy signature waveform $s_k(t)$,
$k = 1, \ldots K$, of the duration T assigned to each of them. The information bits $b_k(i) \in \{\pm 1\}$ for the symbol interval i have the same duration T. The waveform $r(t)$ at the input of the receiver which has a bank of matched filters as its front end, is expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i) \sqrt{a_k} s_k(t - iT - \tau_k) + n(t)$$

where $n(t)$ is a zero-mean, white Gaussian noise with the two-sided power spectral density $N_0/2$, and a_k and τ_k are the received energy and relative delay for user k, respectively. While it is assumed that precise relative delay estimates are available for all users, their amplitudes are considered to be unknown to the receiver. In reality, they are slowly time-varying; however, here they are assumed to remain unchanged over the transmission horizon for each user.

Without a loss of generality, the attention will be on detection of bit 0 of user 1, and it will be assumed that $0 = \tau_1 < \tau_2 < \ldots < \tau_K < T$. The sampled output of the matched filter for user 1 is then:

$$x_1(0) = \sqrt{a_1} b_1(0) + \sum_{k=2}^{K} \sqrt{a_k} \left[\rho_{k1} b_k(-1) + \rho_{1k} b_k(0) \right] + n_1(0)$$

The normalized partial crosscorrelations ρ_{k1} and ρ_{1k}, for $k = 2, \ldots, K$ are:

$$\rho_{k1} = \int_0^T s_1(t)s_k(t+T-\tau_k)dt, \ \rho_{1k} = \int_0^T s_1(t)s_k(t-\tau_k)dt$$

Also, $n_1(0) = \int_0^T n(t)s_1(t)dt$ is a zero-mean Gaussian random variable with variance $N_0/2$ Using the vector notations where $\boldsymbol{\rho}_1 = [\rho_{21}, \ldots, \rho_{K1}, \rho_{12}, \ldots, \rho_{1K}]^T$, $\boldsymbol{b}_1(0) = [b_2(-1), \ldots, b_K(-1), b_2(0), \ \ldots \ , b_K(0)]^T$, $\boldsymbol{A}_1 = diag[\sqrt{a_2}, \ldots, \sqrt{a_K}]$, and $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix}$, the matched filter output is:

$$x_1(0) = \sqrt{a_1}b_1(0) + \boldsymbol{\rho}_1^T \boldsymbol{A}\boldsymbol{b}_1(0) + n_1(0) \tag{1}$$

The multistage detector forms an estimate of the multiuser interference as the weighted vector of tentative decisions on symbols that interfere with $b_1(0)$ directly, and subtracts it from the matched filter output. The final decision statistics $y_1(0)$ and the corresponding final decision for bit 0 of user 1, $\hat{b}_1(0)$, are:

$$y_1(0) = x_1(0) - \boldsymbol{w}_1^T(0)\tilde{\boldsymbol{b}}_1(0), \quad \hat{b}_1(0) = \mathrm{sgn}(y_1(0)) \tag{2}$$

where $\tilde{\boldsymbol{b}}_1(0) = [\tilde{b}_2(-1), \ldots, \tilde{b}_K(-1), \tilde{b}_2(0), \ldots, \tilde{b}_K(0)]^T$ is the vector of tentative decisions affecting bit 0 of user 1, and $\boldsymbol{w}_1(0) = [w_{21}^{(-1)}, \ldots, w_{K1}^{(-1)}, w_{21}^{(0)}, \ldots, w_{K1}^{(0)}]^T$ are the corresponding weights.

Such a detector was first proposed in **[1]** and it utilized fixed weights that required prior estimation of the received signals' amplitudes. Several classes of similar multiuser detectors that use soft-decision tentative statistics were proposed and analyzed in **[2]**.

The multistage detector proposed here operates without the knowledge of the received signals' amplitudes, and without the need to perform their prior estimation in order to set the values of the weights $\boldsymbol{w}_1(0)$. Instead, the weights are determined in an adaptive manner.

II. MULTISTAGE DETECTOR ANALYSIS

In this paper, 1-shot decorrelated tentative decision statistics are considered. This decorrelator, described in **[3]** for a two-user case, can be extended in a straightforward manner for an arbitrary K. The vector of the decorrelator outputs $\boldsymbol{z}_1(0)$ and the corresponding tentative decisions affecting bit 0 of user 1 $\tilde{\boldsymbol{b}}_1(0)$ are:

$$\boldsymbol{z}_1(0) = \boldsymbol{A}\boldsymbol{b}_1(0) + \boldsymbol{\xi}_1, \quad \tilde{\boldsymbol{b}}_1(0) = \mathrm{sgn}(\boldsymbol{z}_1(0)) \tag{3}$$

where $\boldsymbol{\xi}_1 = [\xi_2(-1), \ldots, \xi_K(-1), \xi_2(0), \ldots, \xi_K(0)]^T$ is a zero-mean Gaussian vector having the covariance matrix $\boldsymbol{\Xi}_1$, whose diagonal elements are denoted by $\sigma_{\xi_k}^2$.

For controlling the weights, a steepest descent algorithm which minimizes the output signal power $E\{y_1^2(0)\}$ is used.[3] The weights that achieve this are obtained by iterative

[3]The synchronous version of the detector, and its stability and convergence analysis were presented in **[4]** and **[5]** respectively.

search:

$$\begin{aligned}\boldsymbol{w}_1(1) &\leftarrow \boldsymbol{w}_1(0) - \frac{\mu}{2}\frac{\partial}{\partial \boldsymbol{w}_1(0)}E\{y_1^2(0)\} \\ &= \boldsymbol{w}_1(0) + \mu E\{y_1(0)\tilde{\boldsymbol{b}}_1(0)\} \end{aligned} \tag{4}$$

That is, the algorithm, by minimizing the output power, forces the correlation between the desired output signal and vector of tentative decisions of interfering signals to zero. In practice, the expectation operation is replaced by time averaging.

The steady state values of the weights are evaluated as follows:

$$\begin{aligned} 0 &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{w}_1(0)}E\{y_1^2(0)\} = E\{y_1(0)\tilde{\boldsymbol{b}}_1(0)\} \\ &= E\{x_1(0)\tilde{\boldsymbol{b}}_1(0)\} - E\{\tilde{\boldsymbol{b}}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\}\boldsymbol{w}_1(0) \end{aligned} \tag{5}$$

It is easy to show that $E\{n_1(0)\xi_k(i)\} = 0$, for $k \neq 1$, $i = -1, 0$, so the first of the two expectations above is:

$$E\{x_1(0)\tilde{\boldsymbol{b}}_1(0)\} = \boldsymbol{A}E\{\boldsymbol{b}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\}\boldsymbol{\rho}_1 \tag{6}$$

Clearly, $E\{\boldsymbol{b}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\}$ is diagonal; therefore, the system of $K-1$ linear equations (5), together with (6), gives the steady state values of the weights affecting the first output as:

$$\boldsymbol{w}_1(0) = \left[E\{\tilde{\boldsymbol{b}}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\}\right]^{-1}\boldsymbol{A}E\{\boldsymbol{b}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\}\boldsymbol{\rho}_1 \tag{7}$$

The expressions for the expectations in (7) are shown in the Appendix.

The output error probability is evaluated as follows:

$$\begin{aligned} & P_{e_1} \\ = \; & E_{b_1(0),\boldsymbol{b}_1(0),\tilde{\boldsymbol{b}}_1(0)} Pr\{\hat{b}_1 \text{ in error} | b_1(0), \boldsymbol{b}_1(0), \tilde{\boldsymbol{b}}_1(0)\} \\ = \; & \frac{1}{2}\sum_{\boldsymbol{b}_1(0),\tilde{\boldsymbol{b}}_1(0)} \Big[Pr\{n_1(0) > \sqrt{a_1} - \boldsymbol{\rho}_1^T\boldsymbol{A}\boldsymbol{b}_1(0) \\ + \; & \boldsymbol{w}_1^T(0)\tilde{\boldsymbol{b}}_1(0)\} + Pr\{n_1(0) < -\sqrt{a_1} - \boldsymbol{\rho}_1^T\boldsymbol{A}\boldsymbol{b}_1(0) \\ + \; & \boldsymbol{w}_1^T(0)\tilde{\boldsymbol{b}}_1(0)\}\Big] Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\} Pr\{\boldsymbol{b}_1(0)\} \end{aligned}$$

Since $Pr\{\boldsymbol{b}_1(0)\} = 2^{-(2K-2)}$, and the above expression contains pairwise identical terms (see (A3) in the Appendix), it can be written as:

$$\begin{aligned} P_{e_1} = 2^{-(2K-2)} \sum_{\boldsymbol{b}_1(0),\tilde{\boldsymbol{b}}_1(0)} & [Pr\{n_1(0) > \sqrt{a_1} \\ -\boldsymbol{\rho}_1^T\boldsymbol{A}\boldsymbol{b}_1(0) + \boldsymbol{w}_1^T(0)\tilde{\boldsymbol{b}}_1(0)\}\Big] & Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\} \end{aligned}$$

Finally:

$$P_{e_1} = 2^{-(2K-2)} \sum_{\boldsymbol{b}_1(0),\tilde{\boldsymbol{b}}_1(0)} Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\} \; Q\left(\frac{\sqrt{a_1} - \boldsymbol{\rho}_1^T \boldsymbol{A}\boldsymbol{b}_1(0) + \boldsymbol{w}_1^T(0)\tilde{\boldsymbol{b}}_1(0)}{\sqrt{N_0/2}}\right) \tag{8}$$

where $Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\}$ is the integral of the $(2K-2)$-variate Gaussian density function:

$$Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\} = \frac{1}{\sqrt{(2\pi)^{2K-2}|\boldsymbol{\Xi}_1|}} \int_{\substack{\tilde{b}_k(i)\xi_k(i) = -\sqrt{a_k}b_k(i)\tilde{b}_k(i) \\ k=2,\ldots,K \quad i=-1,0}}^{\infty} \cdots \int exp\left(-\frac{1}{2}\boldsymbol{\xi}_1^T \boldsymbol{\Xi}_1^{-1}\boldsymbol{\xi}_1\right) d\boldsymbol{\xi}_1$$

III. NUMERICAL EXAMPLES AND DISCUSSION

In all the examples, the signal-to-noise ratio for user k is defined as $SNR_k = a_k/N_0$. Figure 1 shows, for two asynchronous users, the probability of error of user 1 versus SNR_1. The relative energy e_1 is defined as $e_1 = \int_{\tau_2}^{T} s_1^2(t)dt$. The case labeled (1) in the figure corresponds to a relatively weak level of interference. It describes a rather unfavorable scenario for this multistage detector due to tentative decisions which are not too reliable. This results in the performance of the detector at higher values of SNR_1 to be somewhat inferior to the decorrelator, whose performance is insensitive to the level of interference. However, the multistage detector outperforms the conventional detector. When the interference is stronger (case (2)), due to the reliable tentative decisions, the multistage detector by far outperforms the other two.

Figure 2 shows the probability of error versus the relative interference energy for the fixed SNR_1. The multistage detector clearly outperforms the decorrelator, and achieves the single-user bound for the relative interference level above 5 dB.

Gold sequences of length 7 [6, Fig. 4] are used as the signature sequences in the next examples. Figure 4 shows the probability of error for user 1 when its energy and the energy of the interferer are the same. The worst case and the average error performance over the values of the relative delay τ_2 are shown. In this scenario the multistage detector outperforms both the conventional and the decorrelating detector, and its average performance, due to the good crosscorrelation properties of the two signature sequences used, is very close to the single user-bound.

Figure 4 shows the error probability of user 1 as a function of relative delays τ_2 and τ_3, with the $SNR's$ of all three users equal to 12 dB. The probability of error of the multistage detector, averaged over all delay pairs is 7×10^{-5}. (The corresponding values for the decorrelator and the conventional receiver are 9.4×10^{-4} and 1.2×10^{-2}, respectively.)

And finally, Figure 5 shows the probability of error versus the relative interference energies for the SNR_1 fixed to 8 dB, and $\tau_2 = T/7$, $\tau_3 = 5T/7$. Again, the multistage detector performs better than the decorrelator and approaches the single user bound as the interferers' energies increase.

IV. CONCLUSION

A two stage multiuser detector is proposed. Its novel feature is the adaptive way in which the multiuser interference estimates are formed. This detector does not require prior estimation of the received signal amplitudes, and is therefore expected to perform better under the channel amplitude varying conditions. The steady state error performance is much better than that of the conventional receiver, and in most cases outperforms the decorrelating detector as well.

V. APPENDIX

The elements on the main diagonal of $E\left\{\tilde{\boldsymbol{b}}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\right\}$ are equal to unity, while the off-diagonal ones are:

$$\begin{aligned} & E\left\{\tilde{b}_k(i)\tilde{b}_l(j)\right\} \\ = \; & E\left\{\operatorname{sgn}\left(\sqrt{a_k}b_k(i)+\xi_k(i)\right)\operatorname{sgn}\left(\sqrt{a_l}b_l(j)+\xi_l(j)\right)\right\} \\ = \; & \frac{1}{2}\sum_{b_k(i),b_l(j)}\left[Pr\{\xi_k(i) < \sqrt{a_k}b_k(i)\,,\,\xi_l(j) < \sqrt{a_l}b_l(j)\right\} \\ & - Pr\{\xi_k(i) < \sqrt{a_k}b_k(i)\;,\;\xi_l(j) > \sqrt{a_l}b_l(j)\}] \\ & i,j = -1,0 \qquad k,l = 2,\ldots,K \quad k \neq l \end{aligned}$$

Each probability term in the above summation defines four integrals. Therefore:

$$E\{\tilde{b}_k(i)\tilde{b}_l(j)\} = \frac{(-1)^m}{2}\sum_{m=1}^{8}\int\int_{D_m} f_{\xi_k\xi_l}d\xi_k d\xi_l \qquad \text{(A1)}$$

where $f_{\xi_k\xi_l}$ denotes the bivariate Gaussian density function of random variables $\xi_k(i)$ and $\xi_l(j)$, and D_m is an appropriate rectangular region of integration.

The diagonal elements of $E\left\{\boldsymbol{b}_1(0)\tilde{\boldsymbol{b}}_1^T(0)\right\}$ are:

$$\begin{aligned} E\{b_k(i)\tilde{b}_k(i)\} \;&=\; 1 - 2Q\left(\sqrt{\frac{a_k}{\sigma_{\xi_k}^2}}\right) \\ & \qquad k = 2,\ldots,K \quad i = -1,0 \qquad \text{(A2)} \end{aligned}$$

with

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$$

By definition:

$$
\begin{aligned}
& Pr\{\tilde{\boldsymbol{b}}_1(0)|\boldsymbol{b}_1(0)\} \\
= \; & Pr\{\mathrm{sgn}\left(\sqrt{a_2}b_2(-1)+\xi_2(-1)\right),\ldots, \\
& \mathrm{sgn}\left(\sqrt{a_K}b_K(-1)+\xi_K(-1)\right),\mathrm{sgn}\left(\sqrt{a_2}b_2(0)\right. \\
& \left.+\xi_2(0)\right),\ldots,\mathrm{sgn}\left(\sqrt{a_K}b_K(0)+\xi_K(0)\right)\} \\
= \; & Pr\{\tilde{b}_2(-1)\xi_2(-1) > -\sqrt{a_2}b_2(-1)\tilde{b}_2(-1),\ldots, \\
& \tilde{b}_K(-1)\xi_K(-1) > -\sqrt{a_K}b_K(-1)\tilde{b}_K(-1), \\
& \tilde{b}_2(0)\xi_2(0) > -\sqrt{a_2}b_2(0)\tilde{b}_2(0),\ldots, \\
& \tilde{b}_K(0)\xi_K(0) > -\sqrt{a_K}b_K(0)\tilde{b}_K(0)\} \\
= \; & Pr\{\tilde{b}_2(-1)\xi_2(-1) < \sqrt{a_2}b_2(-1)\tilde{b}_2(-1),\ldots, \\
& \tilde{b}_K(-1)\xi_K(-1) < \sqrt{a_K}b_K(-1)\tilde{b}_K(-1), \\
& \tilde{b}_2(0)\xi_2(0) < \sqrt{a_2}b_2(0)\tilde{b}_2(0),\ldots, \\
& \tilde{b}_K(0)\xi_K(0) < \sqrt{a_K}b_K(0)\tilde{b}_K(0)\} \\
= \; & Pr\{-\tilde{\boldsymbol{b}}_1(0)|-\boldsymbol{b}_1(0)\} \qquad\qquad (A3)
\end{aligned}
$$

VI. REFERENCES

[1] M.K. Varanasi and B. Aazhang, "Multistage detector in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, No. 4, pp. 509-519, Apr. 1990.

[2] X. Zhang and D. Brady, "Soft-decision multistage detection for asynchronous AWGN channels," *Proc. 31st Annual Allerton Conf. Commun.,* Sept. 1993.

[3] S. Verdú, "Recent progress in multiuser detection," *Advances in Communication and Signal Processing,*
Springer-Verlag, 1989.

[4] Z. Siveski, Y. Bar-Ness and D.W. Chen, "Adaptive multiuser detector for synchronous code division multiple access applications," *1994 Int'l Zurich Seminar Digital Commun.*, ETH Zurich, Switzerland, Mar. 1994, also to appear in *European Trans. Telecommun.*

[5] B. Zhu, N. Ansari, Z. Siveski and Y. Bar-Ness, "Convergence and stability analysis of a synchronous CDMA receiver," accepted at MILCOM'94, and submitted to *IEEE Trans. Commun.*

[6] M.K. Varanasi and B. Aazhang, *"Near-optimum detector in synchronous code-division multiple-access system,"* IEEE Trans. Commun. vol. 39, No. 5, pp. 725-736, May 1991.

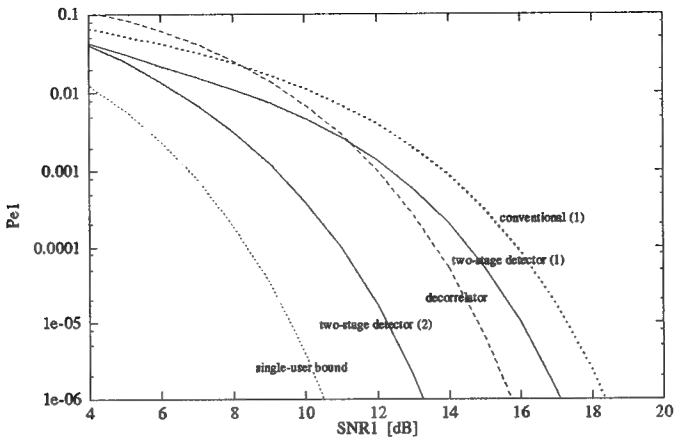


Figure 1: Error probability of user 1 for K=2, $\rho_{12} = 0.2$, $\rho_{21} = 0.6$, $e_1 = 0.4$ (1) $a_2/a_1 = 0.6$ (2) $a_1/a_2 = 0.6$

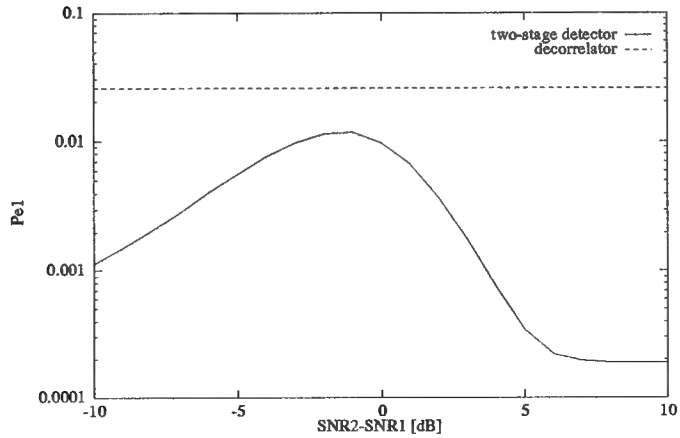


Figure 2: Error probability of user 1 for K=2, $SNR_1 = 8\ dB$, $\rho_{12} = 0.2$, $\rho_{21} = 0.6$, $e_1 = 0.4$

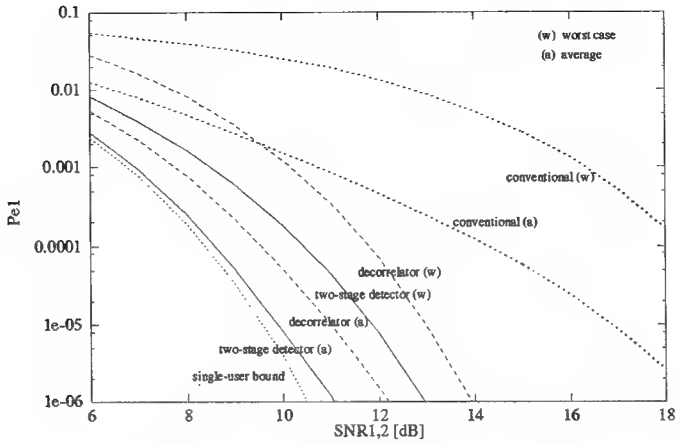


Figure 3: Error probability of user 1 for K=2

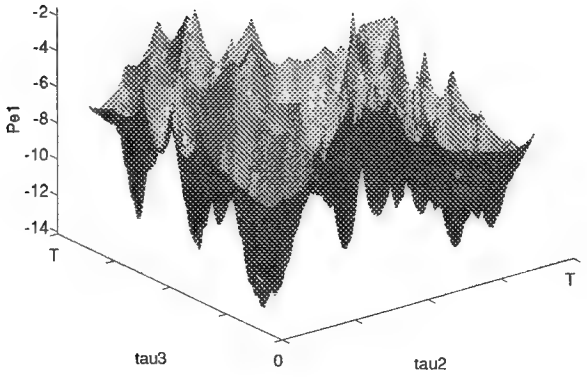


Figure 4: Error probability of user 1 for K=3, $SNR_i = 12\ dB,\ i = 1, 2, 3.$

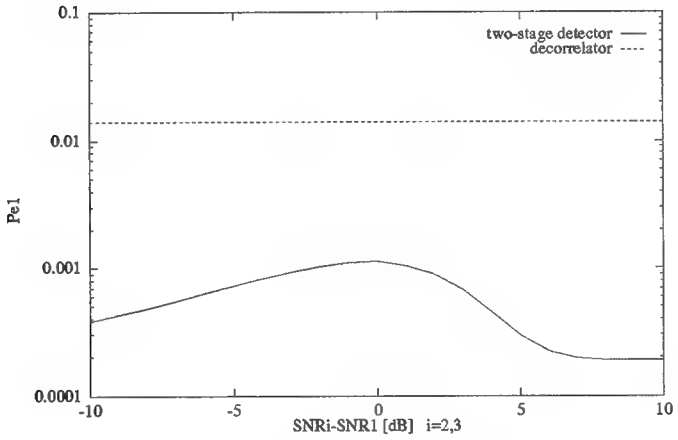


Figure 5: Error probability of user 1 for K=3, $SNR_1 = 8\ dB$

APPENDIX B: PART V

ADAPTIVE MULTIUSER BOOTSTRAPPED DECORRELATING CDMA DETECTOR IN ASYNCHRONOUS UNKNOWN CHANNELS

by
Yeheskel Bar-Ness, David W. Chen and Zoran Siveski

ABSTRACT

For PIMRC uplink multiuser CDMA channels, received signals are time asynchronous, and suffer from amplitude fading. The relative signals' bit delays, energy variations, and hence the cross-correlation matrix of the received signal, are unknown. Using a novel adaptive algorithm, we propose in this paper a decorrelating detector suitable for particular CDMA uplink channels. It is near-far resistant, and can be used by itself if signal-to-noise ratios are high, or as a first stage in the two-stage adaptive canceller previously proposed for the synchronous CDMA case.

I. INTRODUCTION

Multiuser Code Division Multiple Access (CDMA) schemes have become a popular approach for personal, indoor, and mobile wireless radio communications (PIMRC). Conventional single-user detectors implement a bank of filters, each matched to the signature sequence of the particular user. The sampled output of each matched filter contains residual interference from all other users. In some applications, such as military, the bandwidth is not so limited, so that processing gain might be very high and the number of users small enough to result in a low level of controllable interference power. In PIMRC applications bandwidth is scarce and the number of users is large and continually increasing. Particularly for uplink cellular systems, the level of interference can be very large (the near-far problem); communication is asynchronous; and multipath reception is usual.

To handle this difficult, time varying communication environment, multiuser detectors of different structures have been proposed by many researchers. Some [1] are optimal, completely eliminating the near-far problem, although suffering from computational complexity. Others [2,3] only suboptimal, assume the knowledge of the cross correlation matrix of the different user signature sequences and linearly transform the output of the matched filter to obtain uncorrelated signals on which data decisions are made.

Another approach for suboptimum multiuser detectors with low complexity was proposed in [4] and [5]. The performance of some of these suboptimum schemes is close to the performance of the optimum detector. Particularly when the power of the interferers increases, they become indistinguishable. However, these schemes have to perform an estimation of the received signal energies, knowledge of which is required for the detector's proper operation.

At the Center for Communications and Signal Processing, research for improving the performance of multiuser CDMA detectors was directed toward using conventional and novel adaptive algorithms for interference cancellation and signal separation. For example, in [6] and [7], the two-stage multiuser CDMA detector was used in succession with linear transformation, decorrelation, pre-decision (hard limiter), followed by an adaptive interference canceler which used the minimum power criterion. Its error performance was studied and shown to be significantly better than the decorrelating detector. A convergence analysis of this adaptive detector can be found in [8]. Further improvement in performance, particularly at a low interference-to-desired signal ratio, was obtained when the pre-decision was obtained from a soft-limiter [9].

The adaptive interference canceler of the previously mentioned scheme was controlled by a different algorithm called the "bootstrapped decorrelating algorithm" in [10,11,12]. The idea of using this algorithm is to eliminate, in future study, the need for knowing the cross-correlation matrix of the different signals at the output of the matched filter's bank.

Most of the aforementioned multiuser detectors, adaptive or non-adaptive, assume a synchronous data stream of all users and knowledge of the cross-correlation matrix. A suboptimum multiuser detector suitable for asynchronous multiuser CDMA environment was first proposed in [4]. It utilized fixed weights that required prior estimation of the received signals' amplitudes. Several classes of similar multiuser detectors that use soft-decision tentative statistics were proposed and analyzed in [11].

In a companion paper [12] the 1-shot decorrelating tentative decision statistics of Verdu [13] are used. Assuming that the bit-delay overlap between data streams of the different users is known, and using an AWGN non-faded channel, the cross-correlation matrices with zero and one bit-delay can be calculated and used in a minimum power type adaptive canceler similar to the one used with the synchronous channel.

For the case of the faded channel, with unknown energies of the received signal from different users and unknown bit-delays of these signals, the cross-correlation is unknown. For this case, which is typical for uplink PIMRC, we propose in this paper an adaptive decorrelating detector. It might be used as a first stage for the two-stage canceler proposed in [6] and further studied in [7], or by itself if the SNR is high enough so that merely the decorrelating detector, near-far resistant scheme is satisfactory.

After the asynchronous decorrelating model is presented in the next section, the equation of the output for the multiuser case is shown as a function of the adaptive weights. A bootstrapped decorrelating algorithm for controlling these weights is proposed. For the two-user case we find these weights analytically and calculate the corresponding outputs. We find the error probability analytically and depict it as a function of the interference-to-desired signal ratio. Results of some simulations are shown.

II. THE ADAPTIVE DECORRELATOR: MODEL AND ANALYSIS

The proposed asynchronous CDMA adaptive decorrelator consists of a bank of matched filters (matched to the users' signature sequences), followed by a sampler at the bit-rate $1/T$ and controllable weights as shown in Fig. 1.

Let K users share the same channel with the unit energy signature waveform $s_k(t)$, $k = 1, 2, \cdots, K$ of duration T assigned to each one. The received input $r(t)$ can be expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i)\sqrt{a_k}s_k(t - iT - \tau_k) + n(t) \tag{1}$$

where $b_k(i) \in \{+1, -1\}$ are the information bits of the kth user at interval i whose duration is T. $n(t)$ is a zero mean, white Gaussian noise with spectral density $No/2$. a_k and τ_k are the received energy and relative delay for user k. They are unknown to the receiver and may be taken as random variables. The variation of a_k depicts the amplitude fading characteristic of the different users. τ_k may vary with the change of the user's location. In fact each matched filter has only to lock onto the sequence of the corresponding signal. Thereafter will be no need to know accurately the relative delays. The values of these realtive delays will enter into the unknown cross-correlation matrices defined in the sequel. The adaptive algorithm will follow changes in cross correlation matrices caused by variations in relative delays, as well as amplitude variation due to fading. Let us assume that we have ordered the delays such that $-T/2 < \tau_1 < \tau_2 < \cdots < \tau_K < T/2$.

The sampled output of the matched filter for user m at time i (see Figure 2) is:

$$\begin{aligned} x_m(i) &= \int_0^T s_m(t) \left[\sum_{k=1}^{K} b_k(i)\sqrt{a_k}s_k(t - iT - \tau_k) + n(t) \right] \\ &= \sqrt{a_m}b_m(i) + \sum_{k=1}^{m-1} \sqrt{a_k}(\rho_{mk}b_k(i) + \rho_{km}^{(+1)}b_k(i+1) \\ &+ \sum_{k=m+1}^{K} \sqrt{a_k}(\rho_{mk}b_k(i) + \rho_{km}^{(-1)}b_k(i-1) + n_m(i) \end{aligned} \tag{2}$$

where: $\rho_{mk} = \int_0^T s_m(t)s_k(t - iT - \tau_k)dt$

$$\begin{aligned} \rho_{km}^{(+1)} &= \int_0^T s_m(t)s_k(t - (i+1)T - \tau_k)dt \;\; k < m \\ \rho_{km}^{(-1)} &= \int_0^T s_m(t)s_k(t - (i-1)T - \tau_k)dt \;\; k > m \end{aligned} \tag{3}$$

and without loss of generality, we assume zero delay for the corresponding receiver matching code. We also normalize ρ_{mm} so that: $\rho_{mm} = \int_0^T s_m(t)s_m(t - iT - \tau_m)dt = 1$. Also $n_m(i) = \int_0^T n(t)s_m(t)dt$ is a zero-mean Gaussian random variable with variance $N_0/2$.

In matrix notation (2) can be written as:

$$\begin{aligned} \boldsymbol{x}(i) &= \boldsymbol{A}\left[(\boldsymbol{I} + \boldsymbol{\mathcal{P}})\boldsymbol{b}(i) + \bar{\boldsymbol{\mathcal{P}}}_L^T\boldsymbol{b}(i+1) + \bar{\boldsymbol{\mathcal{P}}}_U^T\boldsymbol{b}(i-1)\right] \\ &+ \boldsymbol{n}(i) \end{aligned} \tag{4}$$

where: $\boldsymbol{x}(i) = [x_1(i), x_2(i), \ldots, x_K(i)]^T$

$\boldsymbol{b}(i) = [b_1(i), \ldots, b_K(i)]^T$, $\boldsymbol{A} = diag[\sqrt{a_1}, \cdots, \sqrt{a_K}]$

$$\boldsymbol{\mathcal{P}} = \begin{bmatrix} 0 & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & 0 & \cdots & \rho_{2K} \\ \vdots & \ddots & & \\ \rho_{K1} & \rho_{K2} & \cdots & 0 \end{bmatrix}$$

$$\bar{\boldsymbol{\mathcal{P}}} = \begin{bmatrix} 0 & \rho_{12}^{(+1)} & \cdots & \rho_{1K}^{(+1)} \\ \rho_{21}^{(-1)} & 0 & \cdots & \rho_{2K}^{(+1)} \\ \vdots & \ddots & & \\ \rho_{K1}^{(-1)} & \rho_{K2}^{(-1)} & \cdots & 0 \end{bmatrix}$$

$\bar{\boldsymbol{\mathcal{P}}}_L$ and $\bar{\boldsymbol{\mathcal{P}}}_U$ are the lower and upper triangular matrices of $\bar{\boldsymbol{\mathcal{P}}}$ respectively, and $\boldsymbol{I}$ is the unit matrix.

To generate the decorrelator input $y_m(i)$, we will weigh $x_m(i)$, $m = 1, \cdots, K$ such that:

$$\begin{aligned} y_m(i) &= x_m(i) - \sum_{k=1}^{m-1} (w_{km} x_k(i) + \tilde{w}_{km} x_k(i+1)) \\ &- \sum_{k=m+1}^{K} (w_{km} x_k(i) + \tilde{w}_{km} x_k(i-1)) \end{aligned} \tag{5}$$

or in matrix notation:

$$\boldsymbol{y}(i) = \boldsymbol{x}(i) - \boldsymbol{W}^T \boldsymbol{x}(i) - \tilde{\boldsymbol{W}}_U^T \boldsymbol{x}(i-1) - \tilde{\boldsymbol{W}}_L^T \boldsymbol{x}(i+1) \tag{6}$$

where: $\boldsymbol{y}(i) = [y_1(i), y_2(i), \ldots, y_K(i)]^T$

$$\boldsymbol{W} = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1K} \\ w_{21} & 0 & \cdots & w_{2k} \\ \vdots & \ddots & & \\ w_{K1} & \cdots & & 0 \end{bmatrix}$$

$$\tilde{\boldsymbol{W}} = \begin{bmatrix} 0 & \tilde{w}_{12} & \cdots & \tilde{w}_{1K} \\ \tilde{w}_{21} & 0 & \cdots & \tilde{w}_{2k} \\ \vdots & \ddots & & \\ \tilde{w}_{K1} & \cdots & & 0 \end{bmatrix}$$

and $\boldsymbol{W}_L$ and $\tilde{\boldsymbol{W}}_U$ are the lower and upper triangular matrices of $\tilde{\boldsymbol{W}}$ respectively, after decision:
$\hat{\boldsymbol{b}(i)} = sgn[\boldsymbol{y}(i)]$.

For controlling the weight matrices $\boldsymbol{W}$ and $\tilde{\boldsymbol{W}}$, we use the steepest descent algorithm, which simultaneously reduces the absolute value of the correlation between the decorrelator output $y_m(i)$ and all other outputs of the decorrelator (after decision) at time i, $i-1$ and $i+1$ respectively. The weights w_{mk} and $\tilde{w}_{mk}$ for all $1 \leq k, m \leq K$, $k \neq m$ are controlled by:

$$w_{mk} \leftarrow w_{mk} + \mu E\{y_m(i) sgn(y_k(i))\}$$

$$\tilde{w}_{mk} \leftarrow \tilde{w}_{mk} + \mu E\{y_m(i)sgn(y_k(i-1))\} \quad k > m$$
$$\tilde{w}_{mk} \leftarrow \tilde{w}_{mk} + \mu E\{y_m(i)sgn(y_k(i+1))\} \quad k < m \tag{7}$$

Here we have $2K(K-1)$ weights to be controlled, twice the number we need for the synchronous case (An example of K=2, for finding the steady state values of w_{12}, w_{21}, $\tilde{w}_{12}$, $\tilde{w}_{21}$ will be given in the next section).

In the general case, the signal-to-noise ratios (SNR's) are large enough such that the main contribution to the output errors is due to the multiuser interference. Therefore:

$$\begin{aligned} & E\{y_m(i)\text{sgn}(y_k(i))\} \\ \approx\ & E\{y_m(i)b_k(i)\}(1 - 2Pr(\text{sgn}(y_k(i)) \text{ in error})) \end{aligned} \tag{8}$$

and similarly for the other terms. With these approximations whose validity was examined in [10], we can find the steady state values of w_{mk} and $\tilde{w}_{mk}$ $1 \leq k, m \leq K$, $m \neq k$ by equating $E\{y_m(i)b_k(i)\}$, $E\{y_m(i)b_k(i-1)\}$ and $E\{y_m(i)b_k(i+1)\}$ to zero, respectively.

III. THE TWO-USER CASE

For this case (2) becomes:

$$x_1(i) = \sqrt{a_1}b_1(i) + \rho_{12}\sqrt{a_2}b_2(i) + \rho_{21}\sqrt{a_2}b_2(i-1) + n_1(i)$$
$$x_2(i) = \sqrt{a_2}b_2(i) + \rho_{12}\sqrt{a_1}b_1(i) + \rho_{21}\sqrt{a_1}b_1(i+1) + n_2(i) \tag{9}$$

and the decorrelator output $y_m(i)$, $m = 1, 2$ from (5) becomes,

$$y_1(i) = x_1(i) - w_{12}x_2(i) - \tilde{w}_{21}x_2(i-1)$$
$$y_2(i) = x_2(i) - w_{21}x_1(i) - \tilde{w}_{12}x_1(i+1) \tag{10}$$

A. Algorithm

Correspondingly equations (7) become:

$$w_{12} \leftarrow w_{12} + \mu E\{y_1(i)sgn(y_2(i))\}$$
$$w_{21} \leftarrow w_{21} + \mu E\{y_2(i)sgn(y_1(i))\}$$
$$\tilde{w}_{21} \leftarrow \tilde{w}_{21} + \mu E\{y_2(i)sgn(y_1(i+1))\}$$
$$\tilde{w}_{22} \leftarrow \tilde{w}_{22} + \mu E\{y_2(i)sgn(y_1(i-1))\} \tag{11}$$

These are four non-linear equations that must be solved to get the four different weights. With the approximation of (8) the following four linear equations should be solved instead.

$$E\{y_1(i)b_2(i)\} = 0 \qquad E\{y_2(i)b_1(i)\} = 0$$
$$E\{y_1(i)b_2(i-1)\} = 0 \qquad E\{y_2(i)b_1(i+1)\} = 0 \tag{12}$$

B. Optimal weights

Equation (12) leads to:

$$\begin{aligned}
\rho_{12}\sqrt{a_2} - w_{21}\sqrt{a_2} = 0 &\Rightarrow w_{21} = \rho_{12} \\
\rho_{12}\sqrt{a_1} - w_{12}\sqrt{a_1} = 0 &\Rightarrow w_{12} = \rho_{12} \\
\rho_{21}\sqrt{a_2} - \tilde{w}_{21}\sqrt{a_2} = 0 &\Rightarrow \tilde{w}_{21} = \rho_{21} \\
\rho_{21}\sqrt{a_1} - \tilde{w}_{12}\sqrt{a_1} = 0 &\Rightarrow \tilde{w}_{12} = \rho_{21}
\end{aligned} \tag{13}$$

C. Optimal Outputs

Substituting (13) in (10) together with (9) we get:

$$\begin{aligned}
y_1(i) &= \sqrt{a_1}b_1(i)(1-\rho_{12}^2-\rho_{21}^2) - \rho_{12}\rho_{21}\sqrt{a_1}b_1(i-1) \\
&- \rho_{12}\rho_{21}\sqrt{a_1}b_1(i+1) + \xi_1
\end{aligned} \tag{14}$$

where $\xi_1 = n_1(i) - \rho_{12}n_2(i) - \rho_{21}n_2(i-1)$

Noting that $n_1(i)$, $n_2(i)$ are correlated, we calculate in Appendix A $\sigma_{\xi_1}^2$ as: $\sigma_{\xi_1}^2 = (1 - \rho_{12}^2 - \rho_{21}^2)\frac{N_0}{2}$

The output $y_1(i)$ from eq.(15) depicts the fact that in the steady state the decorrelator output is multiuser interference free, and hence it is near-far resistant. However, it contains inter-symbol interference (ISI). The latter can be eliminated by a suitable equalizer customarily used with fading channels.

D. Probability of Error

The probability of error of detecting $b_1(i)$ from $y_1(i)$ is given (without the equalizer) by (See Appendix B) as:

$$\begin{aligned}
Pe_1 &= \frac{1}{4}\left[Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2-2\rho_{12}\rho_{21})}{\sigma_{\xi_1}}\right)\right. \\
&+ 2Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2)}{\sigma_{\xi_1}}\right) \\
&+ \left.Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2+2\rho_{12}\rho_{21})}{\sigma_{\xi_1}}\right)\right]
\end{aligned} \tag{15}$$

IV. NUMERICAL AND SIMULATION EXAMPLES

Fig. 3 depicts the probability of error of Eq.(15). A constant SNR_1 of 8dB was used, $\rho_{12} = 0.6$, and $\rho_{21} = 0.2$ was assumed in these computations, SNR_2 was varied. For comparison this curve contains the results of simulation with and without equalization. In Fig. 4 we show the error probability of user one when SNR_1 and SNR_2 are varied from $4 - 20$ dB. For comparision the curve contains the result of computation when the cross-correlation is known and linear transformation [13] is used.

V. CONCLUSION

An adaptive multiuser decorrelating detector for asynchronous CDMA was proposed in this paper. The signals' energies and the relative delays between the data streams are assumed unknown. Hence, it is suitable as a model for asynchronous fading PIMRC uplink channels. Using a novel bootstrap blind decorrelating algorithm, it is shown that its steady state leads to a total near-far resistant decorrelation. Computation as well as simulation support this conclusion. Although the two-user case was used as an example, it can be extended to the multiuser case. The decorrelating detector can be used as a first stage for the two-stage canceler previously proposed [6-7] when dealing with uplink PIMRC channels.

VI. Appendix A

$$\begin{aligned}\sigma_{\xi_1}^2 &= E\{n_1^2(i) + \rho_{12}^2 n_2^2(i) + \rho_{21}^2 n_2^2(i-1) \\ &- 2\rho_{12} n_1(i) n_2(i) - 2\rho_{21} n_1(i) n_2(i-1) \\ &+ 2\rho_{12}\rho_{21} n_2(i) n_2(i-1)\}\end{aligned}$$

$E\{n_1(i)n_2(i)\} = \rho_{12}\frac{N_0}{2}$, $E\{n_1(i)n_2(i-1)\} = \rho_{21}\frac{N_0}{2}$, and $E\{n_2(i)n_2(i-1)\} = 0$. Hence:

$$\sigma_{\xi_1}^2 = (1 - \rho_{12}^2 - \rho_{21}^2)\frac{N_0}{2}$$

VII Appendix B

$$\begin{aligned}Pe_1 &= \frac{1}{2}E_{b_1(i-1)b_1(i+1)}\{Pr\{y_1(i) > 0 | b_1(i) = -1\} \\ &+ Pr\{y_1(i) < 0 | b_1(i) = 1\}\} \\ &= E_{b_1(i-1)b_1(i+1)}\{Pr\{\xi_1 > \sqrt{a_1}(1 - \rho_{12}^2 - \rho_{21}^2) \\ &+ \rho_{12}\rho_{21}\sqrt{a_1}b_1(i-1) + \rho_{12}\rho_{21}\sqrt{a_1}b_1(i+1)\}\} \\ &= \frac{1}{4}(Pr\{\xi_1 > \sqrt{a_1}(1 - \rho_{12}^2 - \rho_{21}^2 - 2\rho_{12}\rho_{21})\} \\ &+ 2Pr\{\xi_1 > \sqrt{a_1}(1 - \rho_{12}^2 - \rho_{21}^2)\} \\ &+ Pr\{\xi_1 > \sqrt{a_1}(1 - \rho_{12}^2 - \rho_{21}^2 + 2\rho_{12}\rho_{21})\})\end{aligned}$$

$$= \frac{1}{4}\left(Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2-2\rho_{12}\rho_{21})}{\sigma_{\xi_1}}\right)\right.$$
$$+ 2Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2)}{\sigma_{\xi_1}}\right)$$
$$\left. + Q\left(\frac{\sqrt{a_1}(1-\rho_{12}^2-\rho_{21}^2+2\rho_{12}\rho_{21})}{\sigma_{\xi_1}}\right)\right)$$

VIII. REFERENCES

[1] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple access channels," *IEEE Trans. Inform. Theory,* vol. IT-32, No. 1, pp. 85-96, Jan. 1986.

[2] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code division multiple access channels," *IEEE Trans. Inform. Theory,* vol. IT-35, No. 1, pp. 123-136, Jan. 1989.

[3] R. Lupas and S. Verdu, "Near far resistance of multiuser detectors for synchronous channels," *IEEE Trans.*
Comm., vol. COM-38, No. 4, pp. 496-508, April 1990.

[4] M.K. Varanasi and B. Aazhang, "Multistage detector in asynchronous code division multiple access communications," *IEEE Trans. Comm.,* vol. COM-38, No. 4, pp. 509-519, April 1990.

[5] M.K. Varanasi and B. Aazhang, "Near-optimum detector in synchronous code division multiple access system," *IEEE Trans. Comm.,* vol. COM-39, No. 5, pp. 725 -736, May 1991.

[6] Z. Siveski, Y. Bar-Ness and D.W. Chen, "Adaptive multiuser detector for synchronous code division multiple access applications," *1994 Int'l Zurich Seminar Digital Commun.*, ETH Zurich, Switzerland, Mar. 1994.

[7] Z. Siveski, Y. Bar-Ness and D.W. Chen, "Error performance of synchronous multiuser code division multiple access detector with multidimensional adaptive canceler," accepted for publication in *European Transactions on Telecommunications and Related Technologies.*

[8] B. Zhu, N. Ansari, Z. Siveski and Y. Bar-Ness, "Convergence and stability analysis of a synchronous CDMA receiver," accepted for presentation at *MILCOM' 94* and submitted to *IEEE Trans. Commun.*

[9] D.W. Chen, Z. Siveski and Y. Bar-Ness, "Synchronous multiuser CDMA detector with soft decision adaptive canceler," *28th Annual Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 1994.

[10] Y. Bar-Ness, Z. Siveski and D.W. Chen, "Bootstrapped decorrelating algorithm for adaptive interference cancelation in synchronous CDMA communications systems," *IEEE Third International Symposium on Spread Spectrum Techniques and Applications*, Oulu, Finland, July 1994.

[11] X. Zhang, D. Brady, "Soft-decision multistage detection for asynchronous AWGN channels," *Proc. 31st Annual Allerton Conf. Commun.,* Sept. 1993.

[12] Z. Siveski, L. Zhong and Y. Bar-Ness, "Adaptive Multiuser CDMA Detector for Asynchronous AWGN Channels" *PIMRC'94*, The Hague, The Netherlands, Sept. 1994.

[13] S. Verdú, "Recent progress in multiuser detection,"
Advances in Communication and Signal Processing, Springer-Verlag, 1989.

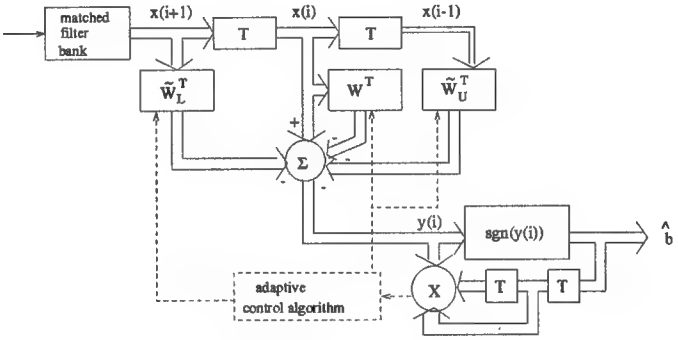


Figure 1: Proposed detector

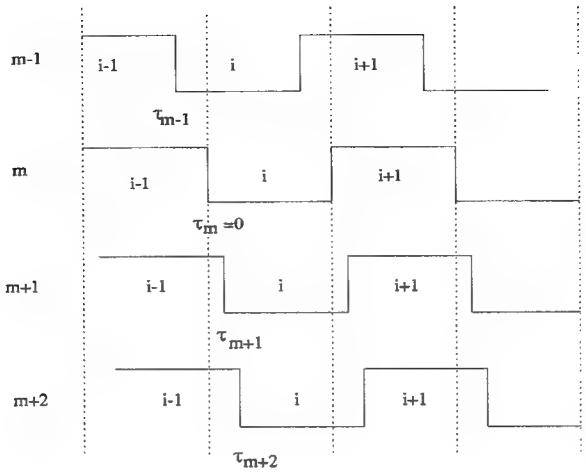


Figure 2: Asynchronous data bits

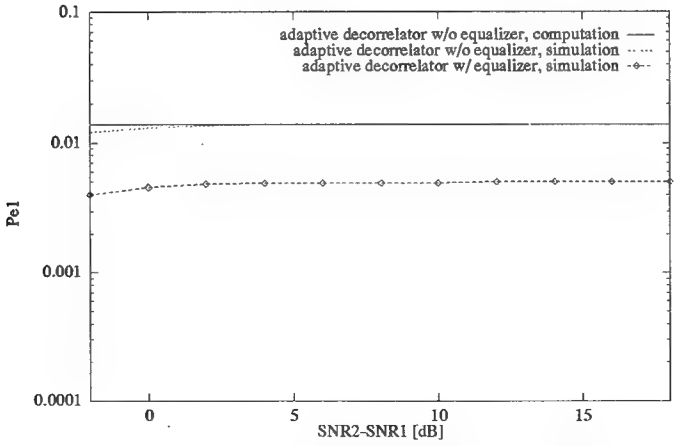


Figure 3: Probability of error of user 1 for $\mathrm{SNR}_1 = 8$ dB, $\rho_{12} = 0.6$, $\rho_{21} = 0.2$

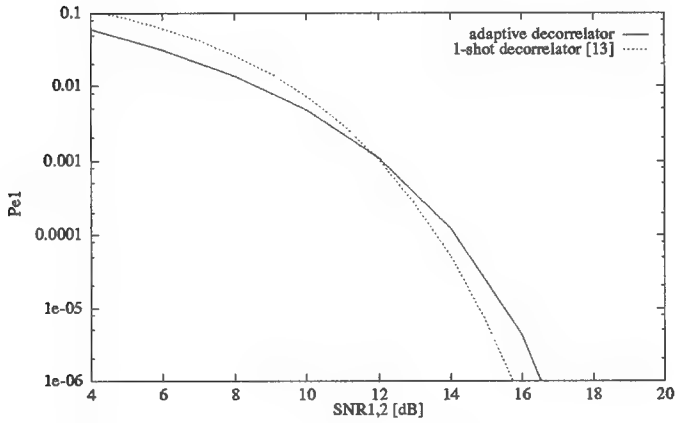


Figure 4: Probability of error of two decorrelators, $\rho_{12} = 0.6$, $\rho_{21} = 0.2$

APPENDIX B: PART VI

CONVERGENCE AND STABILITY ANALYSIS OF A SYNCHRONOUS ADAPTIVE CDMA-BASED PCS RECEIVER

by

Bin Zhu, Nirwan Ansari, Zoran Siveski and Yeheskel Bar-Ness

ABSTRACT:

A thorough investigation on the convergence and stability of a recently proposed adaptive synchronous CDMA receiver is presented in this paper. The receiver consists of a decorrelator at the first stage and an adaptive interference canceler at the second stage. By using a steepest descent algorithm to adaptively control the weights, the knowledge of the users' received powers is no longer required. It was shown that the system has good "near-far" resistance, and can approach optimum performance when the users' SNRs are high. Sufficient conditions for the receiver to achieve convergence are derived, and their properties are analyzed.

I. INTRODUCTION

An optimum receiver in the multiuser environment which performs significantly better than the conventional receiver and also eliminates the infamous near-far problem, albeit at the expense of high computational complexity, was first proposed in [1]. Several suboptimum receivers have been proposed to achieve the reduced computational complexity at the cost of lower performance [2], [3]. The suboptimum two-stage decorrelating receiver analyzed in [3] is one of them. The performance of some of these suboptimum schemes is close to the performance of the optimum detector, particularly when the power of the interferers increases, they become indistinguishable. However, these schemes have to perform an estimation of the received signal energies, knowledge of which is required for the detectors' proper operation. In [4] the detector for synchronous code division multiple access (CDMA) systems that consists of a decorrelator and a nonlinear multidimensional interference canceler whose weights adaptively adjust to the incoming signal was proposed, and its steady state error performance analyzed. It was shown that it provides significantly better performance than the decorrelating detector. This is especially true under the most critical conditions in multiuser environments, such as near-far situations and high bandwidth efficiency utilization. In the presence of strong interference, the proposed detector was shown to achieve the performance of the optimum detector. In the presence of weak interfering signals it performs better than its non-adaptive counterpart in [3], one that requires prior estimation of the received signal

energies.
This paper investigates the learning schemes employed in this adaptive detector and its transient behavior. In Section II, the system model is briefly reviewed and the learning scheme that appears to be most appropriate is proposed and analyzed. Simulation examples that demonstrate the fast convergence of the proposed scheme are presented in Section III.

II. DETECTOR: MODEL AND ANALYSIS

The adaptive synchronous CDMA receiver proposed in [1] consists of a bank of filters (matched to the users' signature sequences) which comprises the front-end, followed by samplers and the decision system, as depicted in Fig. 1.
The received signal $r(t)$ is expressed as:

$$r(t) = \sum_{k=1}^{K} \sum_{i} b_k(i)\sqrt{a_k}s_k(t - iT) + n(t), \tag{1}$$

where $b_k(i) \in \{-1, +1\}$ is the k-th user's data bit in the i-th time interval and $n(t)$ is the additive, zero mean, white Gaussian noise with the two-sided power spectral density $N_0/2$. The received energy of the k-th user signal, unknown to the receiver, is denoted as a_k. The signature sequence $s_k(t)$, of the same duration T as a data bit, is known to the receiver. The sampled outputs of the bank of matched filters in the i-th bit interval can be expressed as:

$$\boldsymbol{x}(i) = \boldsymbol{\mathcal{P}}\boldsymbol{A}\boldsymbol{b}(i) + \boldsymbol{n}(i). \tag{2}$$

For the sake of convenience the index i will be omitted whenever possible in the text. In (2), $\boldsymbol{x} = [x_1, \ldots, x_K]^T$, $\boldsymbol{b} = [b_1, \ldots, b_K]^T$, $\boldsymbol{A} = \text{diag}[\sqrt{a_1}, \ldots, \sqrt{a_K}]$ is a diagonal matrix and $\boldsymbol{n} = [n_1, \ldots, n_K]^T$. The (k, j)th element of the symmetric crosscorrelation matrix $\boldsymbol{\mathcal{P}}$ is defined as:

$$\rho_{kj} = \int_0^T s_k(t)s_j(t)dt \qquad k, j \in (1, 2, \ldots, K)$$

with $\rho_{kk} = 1$. The covariance matrix of a zero mean Gaussian noise vector $\boldsymbol{n}$ is

$$E\{\boldsymbol{n}\boldsymbol{n}^T\} = \frac{N_0}{2}\boldsymbol{\mathcal{P}}$$

The outputs of the decorrelator are given as:

$$\boldsymbol{z} = \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\mathcal{P}}^{-1}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\Gamma}\boldsymbol{n} = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\xi}, \tag{3}$$

where $\boldsymbol{\Gamma}$ is the inverse of $\boldsymbol{\mathcal{P}}$, and $\boldsymbol{\xi} = \boldsymbol{\Gamma}\boldsymbol{n}$. Let $\hat{\boldsymbol{b}} = \left[\hat{b}_1, \ldots, \hat{b}_K\right]^T$ be the decision output of the decorrelator defined as:

$$\hat{\boldsymbol{b}} = \text{sgn}(\boldsymbol{z}). \tag{4}$$

Then the canceler's output is given by:

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{W}^T\hat{\boldsymbol{b}}, \tag{5}$$

where

$$\boldsymbol{W} = \begin{bmatrix} 0 & w_{12} & \dots & w_{1K} \\ w_{21} & 0 & \dots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \dots & 0 \end{bmatrix}$$

The output for the k-th user can be expressed as:

$$y_k = x_k - \boldsymbol{w}_k{}^T \hat{\boldsymbol{b}}_k, \tag{6}$$

where $\boldsymbol{w}_k$ is the k-th column vector of $\boldsymbol{W}$ with the element w_{kk} deleted, and $\hat{\boldsymbol{b}}_k$ is the vector obtained from $\hat{\boldsymbol{b}}$ by deleting the element $\hat{b}_k$. For controlling the weights we use the steepest descent algorithm, which simultaneously minimizes the output signal powers $E\{y_k^2\}$. That is, for the k-th output, the optimum weights are obtained by an iterative search

$$\begin{aligned} \boldsymbol{w}_k(i+1) &= \boldsymbol{w}_k(i) - \frac{\mu}{2}\frac{\partial}{\partial \boldsymbol{w}_k} E\{y_k^2\} \\ &= \boldsymbol{w}_k(i) + \mu E\{x_k \hat{\boldsymbol{b}}_k - \hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T \boldsymbol{w}_k(i)\} \\ &= (\boldsymbol{I} - \mu E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\})\boldsymbol{w}_k(i) + \mu E\{x_k \hat{\boldsymbol{b}}_k\}. \end{aligned} \tag{7}$$

Note that $E\{x_k \hat{\boldsymbol{b}}_k\} = \boldsymbol{w}_k^0 E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\}$. Let $\boldsymbol{c}_k(i+1) = \boldsymbol{w}_k^0 - \boldsymbol{w}_k(i+1)$, then equation (7) can be expressed as:

$$\begin{aligned} \boldsymbol{c}_k(i+1) &= (\boldsymbol{I} - \mu E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\})\boldsymbol{c}_k(i) \\ &= (\boldsymbol{I} - \mu E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\})^{i+1} \boldsymbol{c}_k(0) \\ &= (\boldsymbol{I} - \mu \boldsymbol{R}_k)^{i+1} \boldsymbol{c}_k(0) \\ &= \boldsymbol{H}_k^{i+1} \boldsymbol{c}_k(0), \end{aligned} \tag{8}$$

where both $\boldsymbol{R}_k = E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\}$ and $\boldsymbol{H}_k = \boldsymbol{I} - \mu \boldsymbol{R}_k$ are $(K-1) \times (K-1)$ symmetric matrices. Since $\hat{b}_k \in \{-1, 1\}$, which implies $|E\{\hat{b}_i \hat{b}_j\}| < 1 \quad \forall i \neq j$, the diagonal elements of $\boldsymbol{R}_k$ are equal to 1's and off-diagonal elements range between $(-1, 1)$. Likewise, diagonal elements of $\boldsymbol{H}_k$ are $(1-\mu)$'s, and off-diagonal elements range between $(-\mu, \mu)$. Equation (8) can simply be expressed as:

$$\boldsymbol{c}_k(i) = \boldsymbol{H}_k^i \boldsymbol{c}_k(0). \tag{9}$$

It is easily seen that the training of the weights converges if and only if

$$\lim_{i \to \infty} \boldsymbol{c}_k(i) = 0, \tag{10}$$

which is equivalent to

$$\lim_{i \to \infty} \boldsymbol{H}_k^i = 0. \tag{11}$$

Since $\boldsymbol{H}_k$ is a symmetric matrix, it can be diagonalized and all of its $(K-1)$ eigenvalues are real. So, there always exists an orthogonal matrix $\boldsymbol{Q}_k$ such that

$$\boldsymbol{H}_k = \boldsymbol{Q}_k \boldsymbol{D}_k \boldsymbol{Q}_k^{-1},$$

where $\boldsymbol{D}_k$ is a diagonal matrix $\boldsymbol{D}_k = \text{diag}\,[\lambda_1, \cdots, \lambda_{K-1}]$, and $\lambda_i, i = 1, \cdots, K-1$ are the $(K-1)$ eigenvalues of matrix $\boldsymbol{H}_k$. Thus,

$$\begin{aligned} \boldsymbol{c}_k(i) &= (\boldsymbol{Q}_k \boldsymbol{D}_k^i \boldsymbol{Q}_k^{-1}) \boldsymbol{c}_k(0) \\ &= \boldsymbol{Q}_k \begin{pmatrix} \lambda_1^i & & 0 \\ & \ddots & \\ 0 & & \lambda_{K-1}^i \end{pmatrix} \boldsymbol{Q}_k^{-1} \boldsymbol{c}_k(0). \end{aligned} \tag{12}$$

It is clear that the necessary and sufficient condition for the weights to achieve convergence and stability is

$$| \lambda_i | < 1, \quad \forall i. \tag{13}$$

By applying the Gershgorin theorem [5, p.146] to matrix $\boldsymbol{H}_k$, it is easy to see that matrix $\boldsymbol{H}_k$'s $(K-1)$ real eigenvalues lie in at least one of the $(K-1)$ disks on the complex plane with all their centers reside at $(1-\mu, 0)$ and their radii as r_j $(j = 1, 2, \cdots, K-1)$, where r_j is the sum of the absolute value of each element of either the jth column or the jth row of matrix $\boldsymbol{H}_k$ with the diagonal element deleted:

$$r_j = \sum_{i \neq j} \{\boldsymbol{H}_k\}_{ij} \text{ or } \sum_{\mathrm{i} \neq \mathrm{j}} \{\boldsymbol{H}_{\mathrm{k}}\}_{\mathrm{ji}}. \tag{14}$$

Since the eigenvalues are real values, they can only exist on the real axis and, therefore, all the computations are limited in real domain. The above discussion can also be expressed as:

$$|\lambda_i - (1-\mu)| \leq r_j \quad \forall i, \text{ for some } j. \tag{15}$$

Since the centers of all the disks coincide at $(1-\mu, 0)$, all eigenvalues must lie in the largest disk (the disk with the largest radius). Assume $e_{k_{max}} = max_{i,j}|e_{ij}|$, where e_{ij} is the (i,j)th element of matrix $E\{\hat{\boldsymbol{b}}_k \hat{\boldsymbol{b}}_k^T\}$ (or $\boldsymbol{R}_k$), which depends on the signal SNRs and cross-correlation matrix ρ, then $r_j \leq \mu(K-2)e_{k_{max}}, \forall j$. Therefore, all the eigenvalues must lie in the disk with radius $\mu(K-2)e_{k_{max}}$. That is:

$$(1-\mu) - \mu(K-2)e_{k_{max}} \leq \lambda_i \leq (1-\mu) + \mu(K-2)e_{k_{max}}. \tag{16}$$

Now that the existing range of the eigenvalues of matrix $\boldsymbol{H}_k$ is known, to ensure convergence, the eigenvalues must also lie between $(-1, 1)$. This can be guaranteed if:

1. $(1-\mu) + \mu(K-2)e_{k_{max}} < 1$, which leads to

   $$(K-2)e_{k_{max}} < 1.$$

   This balances three elements of the system: the number of users that can access the system simultaneously, the SNRs, and the cross-correlation of their signature sequences.

2. $(1-\mu) - \mu(K-2)e_{k_{max}} > -1$, which leads to

   $$\mu \leq \frac{2}{1 + (K-2)e_{k_{max}}}.$$

   This is the condition on the learning rate μ for the system to achieve convergence and stability. When condition 1 is satisfied, this condition implies that μ can be any value between $(0, 1)$.

These two conditions are sufficient for the kth signal to converge. It is easy to verify from equation (9) that $\boldsymbol{c}_k(\infty) = 0$ when the above two conditions are satisfied, i.e., the weights will converge to the same steady state no matter what their initial values are.

III. SIMULATION RESULTS AND DISCUSSION

In this section, two sets of Monte-Carlo simulation examples for the transient behavior of the adaptive detector are shown. The first one is a two-user case, while the second one is a three-user case. In either case the SNR of user 1 is fixed at 8 dB, and the SNR(s) of the interferer(s) are varied.

The steady state error performance which was analytically evaluated in [4] is shown in Fig. 2. The crosscorrelation coefficient is equal to 0.7 and represents a high bandwidth efficiency situation. In order to gain some insight into the transient behavior of the system, the adaptation of values of the weights as a function of time is observed. Fig. 3 depicts the average value $\pm$ the standard deviation of w_{21} that is updated according to the rule in (7). Three different SNR_2 values of 2, 8 and 14 dB were considered. The learning rate is ($\mu = 0.2$), and the expected value $E\{x_k\hat{\boldsymbol{b}}_k\}$ is evaluated as a time average, using a sliding window of duration of 2000 steps (signal intervals).

Fig. 4 shows the transient behavior of the corresponding error probabilities for the above scenario. It is observed that the steady state value of the error probability is achieved very quickly.

In the second set of simulation examples, three users having Gold sequences of length seven (frequently used in the literature [3]) are chosen for signature waveforms. The rationale for such a choice is that Gold sequences are regularly used in an asynchronous CDMA environment and the study of its proposed synchronous counterpart may provide a useful indication of the performance of the former. The crosscorrelation matrix $\boldsymbol{\mathcal{P}}$ in this case is:

$$\boldsymbol{\mathcal{P}} = \frac{1}{7}\begin{bmatrix} 7 & -1 & 3 \\ -1 & 7 & -1 \\ 3 & -1 & 7 \end{bmatrix}$$

The analytically evaluated steady state error performance of user 1 is shown in Fig. 5. Fig. 7 describes the updating of the weights w_{21} and w_{31}, for $\mu = 0.2$. Again, as in the two-user case, the steady state values are reached fast, without oscillations afterwards. Fig. 7 shows the probability of error of user 1, the one that is effected by the weights w_{21} and w_{31}. The steady state value is again reached very quickly.

IV. CONCLUSION

It has been demonstrated analytically and by simulation that fast and stable convergence is achievable for the adaptive code division multiple access detector described in [4]. The desirable transient behavior of the learning scheme has enhanced the practicality of the detector, indicating that the steady state performance can be attained in a very fast manner.

V. REFERENCES

[1] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple access channels," *IEEE Trans. Inform. Theory*, Vol. IT-32, No. 1, pp. 85-96, Jan. 1986.

[2] M.K. Varanasi and B. Aazhang, "Multistage detector in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, No. 4, pp. 509-519, Apr. 1990.

[3] M.K. Varanasi and B. Aazhang, "Near-optimum detector in synchronous code-division multiple-access system," *IEEE Trans. Commun.*, vol. 39, No. 5, pp. 725-736, May 1991.

[4] Z. Siveski, Y. Bar-Ness and D. W. Chen, " Error performance of synchronous multiuser code division multiple access detector with multidimensional adaptive canceler," accepted for publication in *European Trans. on Telecommunications*, also presented at *1994 Int'l Zurich Seminar Digital Comm.*, pp. 67-75.

[5] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequality*, Allyn and Bacon, Inc., Boston, 1964.

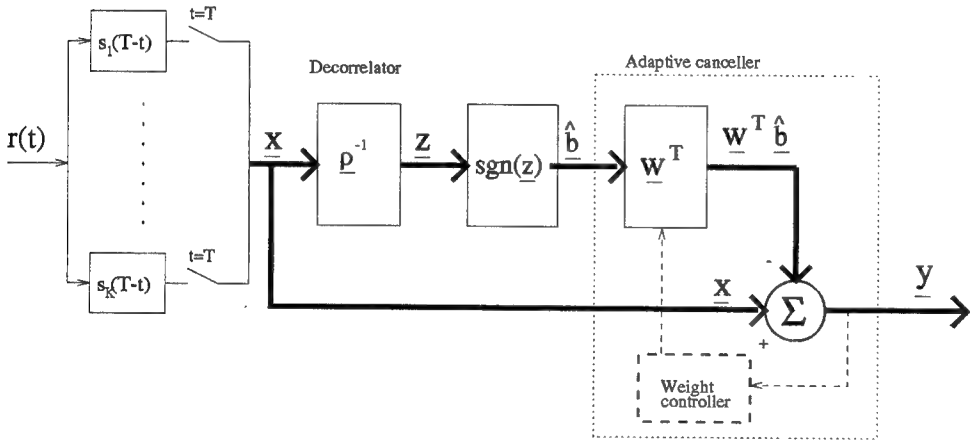


Figure 1: The proposed receiver scheme

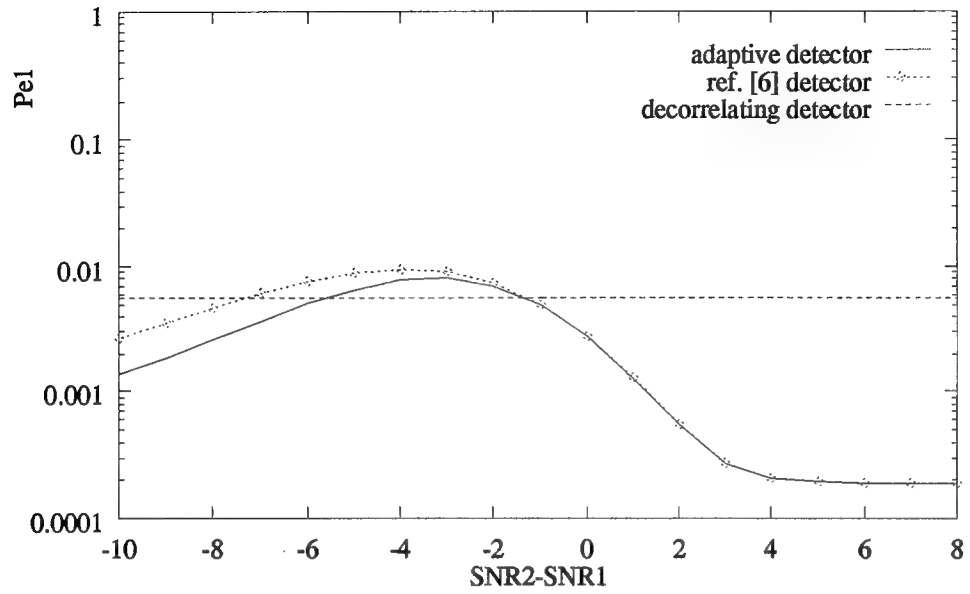


Figure 2: Error probability of user 1 in the two-user case with $\rho_{12} = 0.7, SNR_1 = 8dB$

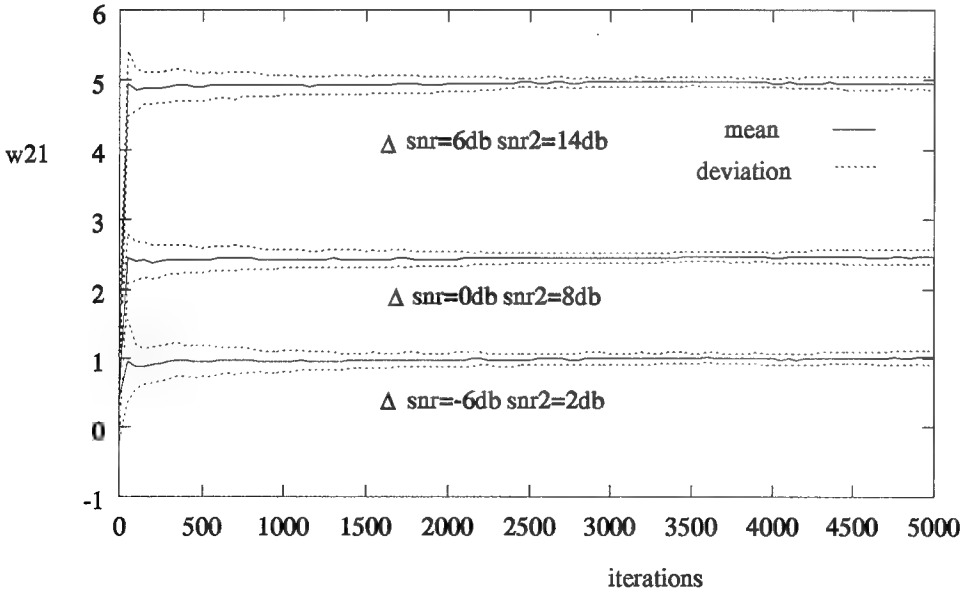


Figure 3: Convergence of w_{21} in the two-user case using the proposed learning scheme (7) with $\mu = 0.2, \rho_{12} = 0.7$

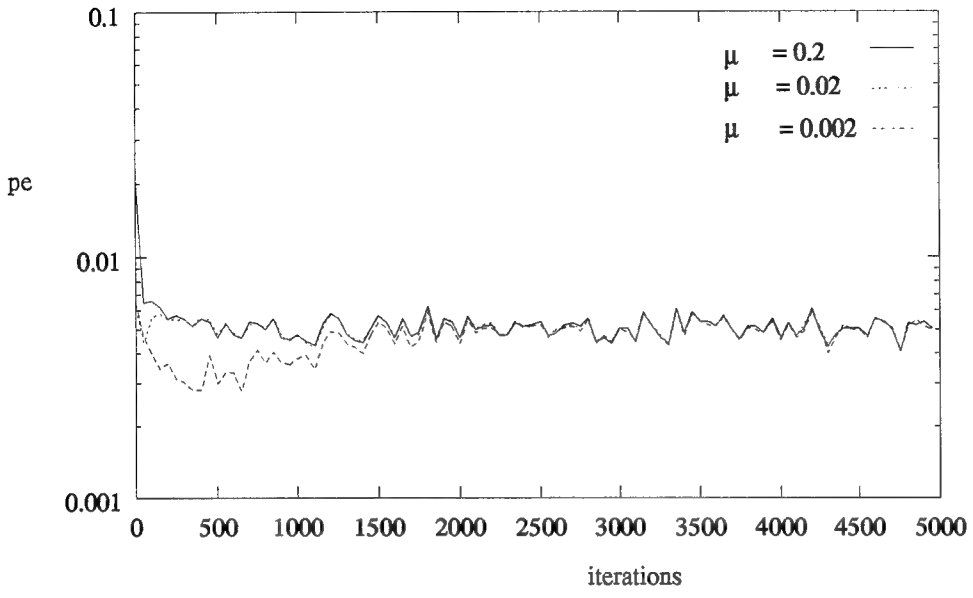


Figure 4: Error probability of user 1 in the two-user case with $SNR_1 = 8dB, SNR_2 = 2dB, \rho_{12} = 0.7$

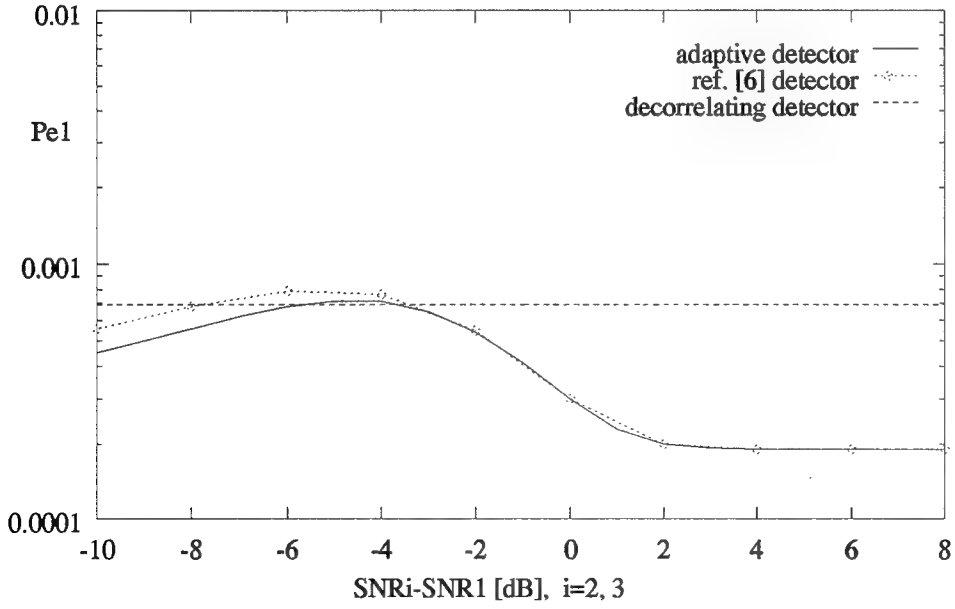


Figure 5: Steady state error probability of user 1 in the three-user case with $SNR_1 = 8dB$

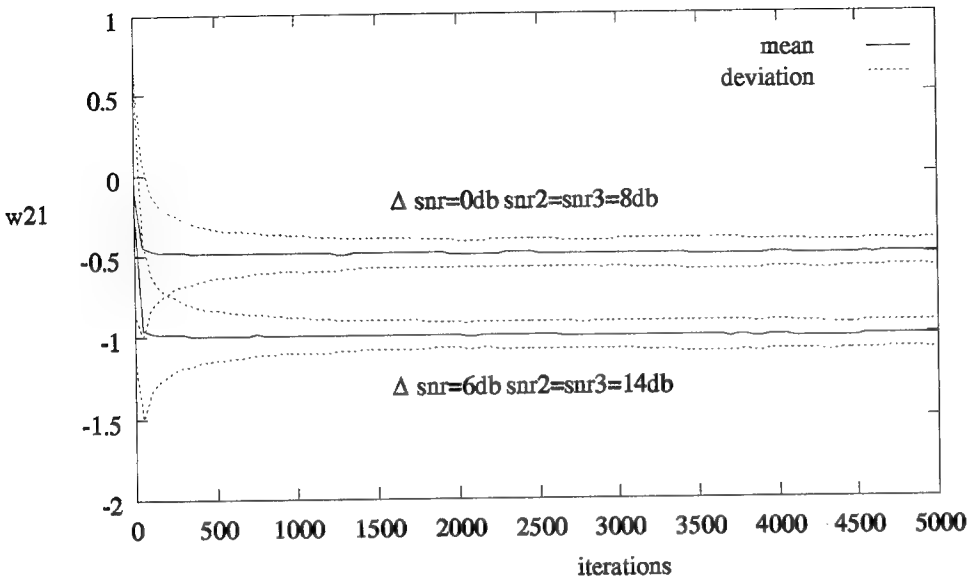


Figure 6: Convergence of w_{21} in the three-user case using the proposed learning scheme (9) with $\mu = 0.2$

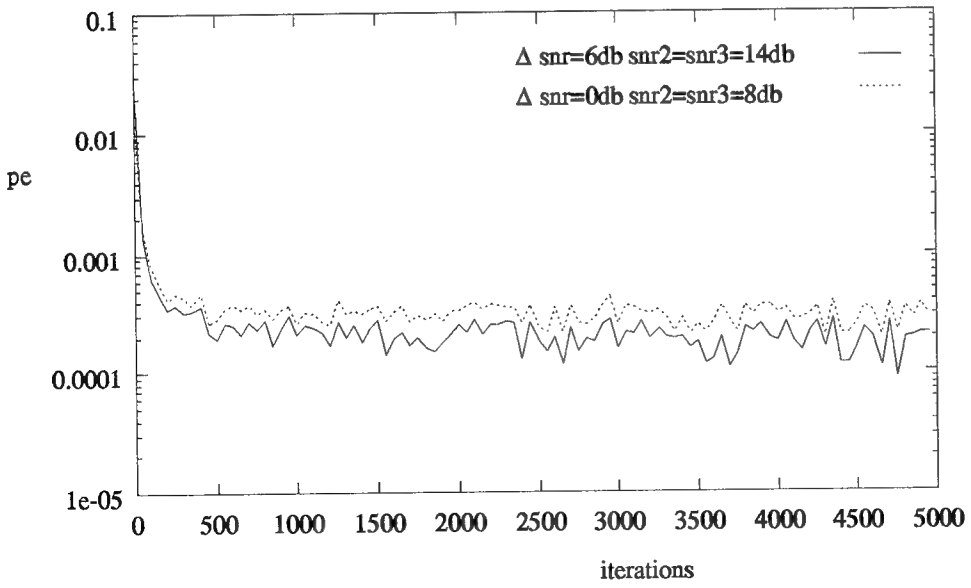


Figure 7: Error probability of user 1 in the three-user case with $\mu = 0.2$

APPENDIX C: PART I

BLIND DECISION FEEDBACK EQUALIZATION BASED ON DECORRELATION

by
Raafat Kamel and Yeheskel Bar-Ness

ABSTRACT

A new blind equalization algorithm is presented which is based on decorrelating the equalizer's output. The algorithm is used with a decision feedback equalizer. The performance of the algorithm is studied analytically and through computer simulation. The algorithm is shown to converge to the correct channel parameters irrespect of the initial error rate. Upper and lower bounds on the probability of error were derived for the decision feedback equalizer. While the upper bound has been known for a long time, the lower bound is presented here for the first time. It is shown that the lower is tighter than the commonly used "matched filter" or "no ISI" bound. A Kalman-type decorrelation algorithm for blind decision feedback equalization is also introduced, which has a higher convergence rate than the original decorrelation algorithm.

I. INTRODUCTION

In digital communication the need for blind equalization arises in when the channel is unknown to the receiver, which attempts to estimate a sequence of transmitted data without resorting to the use of a training sequence. The main problem of blind equalization is that of finding an appropriate cost function (or equivalent error function) that reflects the amount of intersymbol intersymbol (ISI) introduced by the channel, and which does not involve the transmitted symbols [1-5]. Optimization of the cost function should lead to minimization of the ISI. In other words, optimization of such a function should be consistent with the minimization of ISI.

The first known blind equalization algorithm was introduced by Sato [1]. Sato's blind equalization algorithm error function was later generalized by Benveniste, *et al.,* [3]. Considering Sato's cost function, Godard [2] then described a class of cost functions that generalizes Sato's. It is also worth mentioning that the constant modulus algorithm (CMA) developed separately by Treichler, *et al.,* [6-7] is also a special case of Godard's algorithm.

Bellini, *et al.* [4], followed a different approach and developed what they termed "Bussgang Techniques." Based on some assumptions about the equalizer and the channel parameters, they derived a maximum likelihood estimator of the reference signal. This estimator depends on the type of modulation used and the signal-to-noise ratio (SNR). The aforementioned algorithms are viewed as special cases of the Bussgang technique.

The above algorithms used non-convex cost functions which, in turn, imply the existence of local minima to which the blind equalizer might converge. Some of these equilibria may be undesirable, *i.e.*, at those equilibria the equalizer will not be able to remove ISI. This was shown and demonstrated by Ding, *et al.,* for the Godard algorithm [8-9] and for the Sato algorithm [10]. In [11], the ill convergence of the Benveniste, *et al.*, algorithms [3] was also considered, thus proving that none of the previous algorithms were globally convergent. For these algorithms equalizer initialization becomes an important issue. One would initialize the equalizer away from the neighborhood of the ill-convergent minima.

Verdu, *et al.*, [12] developed a technique that insures global convergence of blind equalizers. The key observation in [12] is that over-parameterizing the blind equalizer is the prime cause of ill-convergence, and hence anchoring (setting the first coefficient to one) the blind equalizer was proposed [12][13]. This, together with using a convex function, guarantees convergence. Verdu used the minimized energy as a cost function. Vembu, *et al.* [13], used the l_1 norm of the equalizer weight as a cost function, which was approximated by the l_p norm of the equalizer output. Kennedy and Ding [14] applied the concept of anchored equalization to a QAM transmission. Due to the complex nature of the signal constellation, they performed joint equalization and carrier recovery. This was done by anchoring the sum of the real and imaginary parts of the center tap to 1, and using the maximum of either the real or the imaginary part of the equalizer output as a cost function. The cost function was implemented using the l_p norm of the real or imaginary part of the equalizer's output.

Another family of blind equalization algorithms that appears in the literature is based on high-order moments and polyspectra [15] and [16]. In general, these algorithms give better performance at the expense of higher arithmetic complexity. Basically, they use the received samples to estimate the channel parameters and reconstruct the transmitted data via inverse filtering. The computational complexity of these algorithms makes them inappropriate for on-line processing.

A different approach to blind equalization is based on maximum likelihood sequence estimation [17-19]. In [17], the channel is estimated using the least squares method. In [18] an iterative procedure was devised which processes a frame of received data. A least square technique followed by a Viterbi algorithm (VA) is iterated for jointly estimating the channel and data sequence. A similar iterative procedure is used in [19], except the channel is estimated using the EM algorithm.

The main emphasis of blind equalization has been on the linear equalizer structure. We, on the other hand, address another structure which receives less attention: the decision feedback equalizer. In this paper we introduce a new blind equalization criterion which will be used in conjunction with a decision feedback equalizer.

In this paper, we assume that the original data is independent and identically distributed (iid). This is a valid and widely used model. At the output of the channel, the data is no longer independent. The channel introduces the correlation in the form of the ISI. We exploit the white noise-like characteristics of the original signal and adapt the blind equalizer using decorrelation. This was motivated by [20] who provided a simple test to show that an adaptive equalizer has converged to the correct settings. In [20] it was shown that if the input data is binary and iid, then the

decorrelation at the output of the slicer of a decision directed equalizer is a necessary and sufficient condition for the correct convergence of the equalizer. In this paper we show that the decorrelation at the input of a slicer is a necessary and sufficient condition for the perfect cancellation of ISI. Hence we use decorrelation as a criterion for equalization.

This paper is organized as follows. In section 2, we introduce the decorrelation algorithm for blind equalization. In section 3 we discuss the equalization of autoregressive channels (AR). In section 4 the decorrelation algorithm is used with the decision feedback equalizer to equalize a channel of the finite impulse response type. In section 5 upper and lower bounds on the probability of error of the decision feedback equalizer in the presence of additive white Gaussian noise (AWGN) are given. Due to the simplicity of the decorrelation algorithm one could derive a Kalman-type decorrelation blind equalizer, which increases the convergence substantially. This is presented in section 6. Conclusions are given in section 7.

II. THE DECORRELATION ALGORITHM

Consider an $N-$tap adaptive equalizer, whose weights are denoted by w_i, $1 \leq i \leq N$. These weights are updated using the steepest decent method, whose update equation can be expressed as

$$w_i^{k+1} = w_i^k - \mu f(\cdot) \text{ for } i = 1, \cdots, N, \tag{1}$$

where μ is the constant of adaptation and $f(\cdot)$ is the error function of the algorithm. The roots of the error function determine the steady state of the algorithm.

In this paper we propose a decorrelation algorithm which uses $\overline{A_k A_{k-i}}$ as an error function, where A_k is the output of the equalizer in the linear equalizer case and the input to the slicer in the decision feedback equalizer (DFE) case. Therefore the weight update equation can be written as

$$w_i^{k+1} = w_i^k + \mu \overline{A_k A_{k-i}} \text{ for } i = 1, \cdots, N.$$

In a practical implementation one would replace the expectation by the current realization, leading to the stochastic difference equation,

$$w_i^{k+1} = w_i^k + \mu A_k A_{k-i} \text{ for } i = 1, \cdots, N. \tag{2}$$

The mean steady state equalizer weight is found by setting the error function in equation (1) (in our case $\overline{A_k A_{k-i}}$ for $i = 1, \cdots, n$) to zero. In the following sections we consider different channel models and show that the decorrelation algorithm converges in the mean to the channel parameters and hence cancels ISI.

III. EQUALIZATION OF AUTOREGRESSIVE CHANNELS

In this section we introduce the decorrelation criterion for equalizing an AR channel. Despite the fact that the AR model is not commonly used as a channel model for

equalization purpose, it gives a good insight to the decorrelation algorithm. Although our main emphasis in this paper is the decision feedback equalizer, we use the linear equalizer in this section to motivate the decorrelation algorithm presented in the previous section.

3.1 Problem Formulation

Consider an AR(n) channel driven by an equi-probable binary sequence $\{I_k\}$. The output X_k is given by

$$X_k = gI_k + \sum_{i=1}^{n} \alpha_i X_{k-i}, \tag{3}$$

where g is the channel gain and α's are the channel parameters.

Figure 1 shows the anchored FIR blind equalizer. In the previous section, we mentioned the advantages of using an anchored equalizer [12]. The output of the anchored equalizer is given by

$$\begin{aligned} A_k &= X_k - \sum_{i=1}^{n} w_i X_{k-i} \\ &= gI_k + \sum_{i=1}^{n} (\alpha_i - w_i) X_{k-i}. \end{aligned}$$

The above equation can be written as

$$A_k = g\left(I_k - \sum_{i=1}^{n} w_i I_{k-i}\right) + \sum_{i=1}^{n} \alpha_i A_{k-i}. \tag{4}$$

In order to show that the equalizer converges in the mean to channel parameters, we first derive some correlation results. Multiply equation (4) by $A_{k-(n+1)}$ and take the expectation, to get

$$\begin{aligned} \overline{A_k A_{k-(n+1)}} &= g\left(\overline{I_k A_{k-(n+1)}} - \sum_{i=1}^{n} w_i \overline{I_{k-i} A_{k-(n+1)}}\right) + \sum_{i=1}^{n} \alpha_i \overline{A_{k-i} A_{k-(n+1)}} \\ &= \sum_{i=1}^{n} \alpha_i \overline{A_{k-i} A_{k-(n+1)}}. \end{aligned} \tag{5}$$

The last step follows from the fact that $I_{k-i} A_{k-(n+1)}$ are independent for $i = 0, 1, 2, \cdots, n$. It further follows that

$$\begin{aligned} \overline{A_k A_{k-(n+1)}} &= \sum_{i=1}^{n} \alpha_{n-i+1} \overline{A_k A_{k-i}}, \\ &= 0 \end{aligned}$$

since we require that $\overline{A_k A_{k-i}} = 0$ for $i = 1, 2, \cdots, n$ at the steady state. This proves that $\overline{A_k A_{k-i}} = 0$ for $i = n+1$. Similarly, one can prove that $\overline{A_k A_{k-i}} = 0$ for $i > n$. Therefore, at steady state we have

$$\overline{A_k A_{k-i}} = \sigma_A^2 \delta(i), \tag{6}$$

where $\delta(\cdot)$ is the Kronecker delta, and we assumed that A_k is a wide sense stationary random process. In other words, the sequence $\{A_k\}$ becomes white noise-like.

3.2 Steady State Analysis

In the Z-domain, one may write equation (4) as

$$\frac{1-\alpha_1 z^{-1}-\cdots-\alpha_n z^{-n}}{1-w_1 z^{-1}-\cdots-w_n z^{-n}}A(z) = gI(z),$$

where $A(z)$ and $I(z)$ are the Z-transforms of A_k and I_k, respectively. In the time domain, using long division, the above equation may be written as

$$\sum_{i=0}^{\infty}\gamma_i A_{k-i} = gI_k, \tag{7}$$

where

$$\begin{aligned}
\gamma_0 &= 1\\
\gamma_1 &= w_1-\alpha_1\\
\gamma_2 &= (w_2-\alpha_2)-w_1\gamma_1\\
\gamma_3 &= (w_3-\alpha_3)-w_2\gamma_1-w_1\gamma_2\\
&\vdots\\
\gamma_n &= (w_n-\alpha_n)-w_{n-1}\gamma_1-w_1\gamma_{n-1}\\
&\vdots
\end{aligned}$$

Now multiply equation (7) by A_{k-1} and take expectation

$$\sum_{i=0}^{\infty}\gamma_i\overline{A_{k-i}A_{k-1}} = 0.$$

The above result follows since A_{k-1} is independent of I_k. Hence together with equation (6), we have

$$\sigma_A^2\gamma_1 = 0;$$

therefore, we get $\gamma_1 = 0$, *i.e.*, $w_1 = \alpha_1$. Similarly, one can show that $\gamma_2 = 0$, which together with $\gamma_1 = 0$, gives $w_2 = \alpha_2$. Hence

$$w_i = \alpha_i \qquad \text{for } i = 1, 2, \cdots, n.$$

Therefore, at steady state, the decorrelation algorithm results in perfect ISI cancellation.

IV. EQUALIZATION OF MOVING AVERAGE CHANNELS

In this section we discuss the blind decision feedback equalization of channels with a finite impulse response. We will assume that the channel will not introduce any precursors, and as a result the DFE will include a feedback filter and a decision device. A typical DFE will include a forward filter to cancel the precursor ISI. This filter is typically a linear equalizer followed by a linear prediction error filter. Since many articles deal with blind linear equalization we chose not to address it. In a real scenario one could use the decorrelation algorithm to adapt the feedback filter and any of the existing blind algorithms for the forward filter.

4.1 Problem Formulation

The channel and equalizer model under consideration is shown in Figure 2. The cascade of transmit, channel and receive filters is modeled as an FIR filter with impulse response

$$h(n) = 1 + \sum_{i=1}^{N} h_i \delta(n-i),$$

where $\delta(\cdot)$ is the Kronecker delta. In the above equation we normalized relative to the first cursor (h_0). We also assume that the input I_k is a binary white sequence with a zero mean. The output of the channel is thus given by

$$X_k = I_k + \sum_{i=1}^{N} h_i I_{k-i}.$$

We assume that the channel is slowly time varying and the receiver has perfect carrier and timing recovery. The channel post-cursors $\{h_1, \cdots, h_N\}$ introduce ISI on the current data symbol I_k. The estimated data $\hat{A}_k$ is produced by passing A_k through a slicer.

Referring to Figure 2, the input to the slicer of the decision feedback equalizer A_k is given by

$$\begin{aligned} A_k &= X_k - \hat{\mathbf{A}}'_{k-1}\mathbf{W} \\ &= I_k + \mathbf{I}'_{k-1}\mathbf{H} - \hat{\mathbf{A}}'_{k-1}\mathbf{W}, \end{aligned} \tag{8}$$

where $\hat{\mathbf{A}}_{k-1}$ is the vector of the past N decisions $\hat{\mathbf{A}}'_{k-1} = [\hat{A}_{k-1}, \hat{A}_{k-2}, \cdots, \hat{A}_{k-N}]$ (the prime stands for transpose) and $\mathbf{I}_{k-1}$ is the vector of past transmitted information bits $\mathbf{I}'_{k-1} = [I_{k-1}, I_{k-2}, \cdots, I_{k-N}]$, where $I_{k-i} \in \{-1, 1\}$ and $P\{I_{k-i} = 1\} = P\{I_{k-i} = -1\} = \frac{1}{2}$. $\mathbf{W}$ and $\mathbf{H}$ are the equalizer and channel parameter vectors, respectively;$\mathbf{W}' = [w_1, w_2, \cdots, w_N]$ and $\mathbf{H}' = [h_1, h_2, \cdots, h_N]$.

In this paper we will assume a noiseless situation, *i.e.*, we consider an arbitrary high signal-to-noise ratio. AWGN will be considered in the next section For ideal ISI cancellation, the slicer's input $A_k = I_k$ and therefore sequence $\{A_k\}$ will be decorrelated, *i.e.*, $\overline{A_k A_{k-n}} = 0$ for $n \neq 0$. In other words, decorrelation is a necessary condition for ideal cancellation of ISI. In order to be able to use the decorrelation of the slicer's input as a criterion for controlling the feedback weight vector $\mathbf{W}$, we must prove that decorrelation is also sufficient for cancelling ISI. This is what we intend to show in the next section.

4.2 Sufficiency

In order to prove sufficiency, we rewrite equation (8) as

$$A_k = I_k + \sum_{i=1}^{N} h_i I_{k-i} - \sum_{i=1}^{N} w_i \hat{A}_{k-i}. \tag{9}$$

If we denote the set of all correct decisions by A_1 and the set of all incorrect decisions by A_2, *i.e.*,

$$\begin{aligned} A_1 &= \{\hat{A}_i : \hat{A}_i = I_i\} \\ A_2 &= \{\hat{A}_i : \hat{A}_i = -I_i\}, \end{aligned}$$

then equation (9) can be written as

$$\begin{aligned} A_k &= I_k + \sum_{i:\hat{A}_{k-i} \in A_1} (h_i - w_i) I_{k-i} + \sum_{i:\hat{A}_{k-i} \in A_2} (h_i + w_i) I_{k-i} \\ &= I_k + \sum_{i=1}^{N} \gamma_i I_{k-i} \\ &= I_k + isi_k, \end{aligned} \tag{10}$$

where γ_i is given by

$$\gamma_i = \begin{cases} (h_i - w_i) & \text{for all } i : \hat{A}_{k-i} \in A_1 \\ (h_i + w_i) & \text{for all } i : \hat{A}_{k-i} \in A_2, \end{cases}$$

and

$$isi_k = \sum_{i=1}^{N} \gamma_i I_{k-i}.$$

We can now show that decorrelation is a sufficient condition for cancelling ISI. Multiply equation (10) by $\mathbf{A}_{k-1} = [A_{k-1}, A_{k-2}, \cdots, A_{k-N}]$, the vector of the past slicer's input, to obtain

$$A_k \mathbf{A}_{k-1} = I_k \mathbf{A}_{k-1} + \sum_{i=1}^{N} \gamma_i I_{k-i} \mathbf{A}_{k-1}.$$

By taking the expectation on both sides of the above equation, it can be shown that the above equation reduces to

$$\begin{pmatrix} \overline{A_k A_{k-1}} \\ \overline{A_k A_{k-2}} \\ \vdots \\ \overline{A_k A_{k-n}} \\ \vdots \\ \overline{A_k A_{k-N+1}} \\ \overline{A_k A_{k-N}} \end{pmatrix} = \begin{pmatrix} \gamma_1 + \sum_{i=1}^{N-1} \gamma_i \gamma_{i+1} \\ \gamma_2 + \sum_{i=1}^{N-2} \gamma_i \gamma_{i+2} \\ \vdots \\ \gamma_n + \sum_{i=1}^{N-n} \gamma_i \gamma_{i+n} \\ \vdots \\ \gamma_{N-1} + \gamma_1 \gamma_N \\ \gamma_N \end{pmatrix}. \tag{11}$$

It is clear from the last entry of the vectors in the above equation that $\gamma_N = 0$ iff $\overline{A_k A_{k-N}} = 0$. Similarly, it follows from the $(N-1)$th entry that if $\gamma_N = 0$, then $\gamma_{N-1} = 0$ iff $\overline{A_k A_{k-N+1}} = 0$. One would thus start from the bottom entry and use back substitution to show that $\gamma_i = 0$ for $i = 1, \cdots, N$ iff $\overline{A_k A_{k-i}} = 0$ for $i = 1, \cdots, N$. It thus follows from equation (10) that $isi_k = 0$ in the steady state iff $\overline{A_k A_{k-i}} = 0$ for $i = 1, \cdots, N$. This completes the proof that decorrelation is also sufficient for cancelling ISI.

4.3 Simulation Results

A non-minimum phase channel, whose transfer function is given by

$$H(z^{-1}) = 1 + 0.8z^{-1} + 0.4z^{-2} - 0.6z^{-3} + 0.2z^{-4},$$

was used in a simulation to demonstrate the convergence of the decorrelation algorithm. Figure 3 shows the learning curve obtained by ensemble-averaging over 200 independent simulation runs. The blind DFE used had four tap gains in the feedback filter. The adaptation constant μ used in the simulation was set to 0.01, and the weights were initialized to zero. As depicted in Figure 3, the four weights $w_i, i = 1, 2, 3, 4$ converge to the correct channel parameters.

The decorrelation algorithm was also used to equalize a non-minimum phase channel whose transfer function is given by

$$H(z^{-1}) = 1 + 0.5z^{-1} - 1.44z^{-2}.$$

In order the demonstrate the ability of the decorrelation algorithm to converge to the correct weight irrespective of the initial condition, we initialized the equalizer weights with different values. Figure 4 portrays the trajectories for different equalizer initializations. It clearly shows that the decorrelation algorithm always converges to the correct weights.

V. ERROR ANALYSIS

In considering the probability of bit error for the AWGN case with a zero mean and a variance of σ^2, we follow the Duttweiler, *et al.*, approach [21]. Only results will be given in this section, readers interested in details are referred to [21] [22]. An upper bound to the probability of error was derived in [21]. Duttweiler, *et al.,* modeled the errors in the feedback filter of a DFE as a finite state machine. Using the state transition probabilities ($\alpha_i, \quad i = 1, \cdots, N$ where in our case N is the number of delay elements in the feedback filter), Duttweiler obtained an upper bound to the probability of error. Using the same finite state machine, a lower bound was derived in [22] which was tighter than the commonly used "no ISI" bound.

It was shown in [21] that the probability of error q is given by

$$q \;=\; R_N^{-1},$$

where

$$R_N \;=\; 1 + \sum_{i=0}^{N-2} \prod_{m=0}^{i} \alpha_m + (1 - \alpha_N)^{-1} \prod_{m=0}^{N-1} \alpha_m. \tag{12}$$

In order to calculate the probability of error, one has to know all the state transition probabilities, α_i. This is not feasible for large N. Instead, one can use lower bounds on the transition probability to derive an upper bound on the probability of error.

In [22] we derived expressions for the lower and upper bounds on the transition probabilities, then obtained respectively an upper and lower bound on the probability of error of the blind decision feedback equalizer.

The bounds on the transition probabilities are given below (details of the derivation are omitted).

$$\begin{aligned}
\alpha_m &\geq \frac{1}{2}\left(Q\left(\frac{-1-2h_{m+1}+\beta_{m+1}}{\sigma}\right)+Q\left(\frac{-1+2h_{m+1}+\beta_{m+1}}{\sigma}\right)\right), \\
&\qquad \text{for } 0 \leq m \leq N-2 \\
\alpha_m &\leq \frac{1}{2}\left(Q\left(\frac{-1-2h_{m+1}-\beta_{m+1}}{\sigma}\right)+Q\left(\frac{-1+2h_{m+1}-\beta_{m+1}}{\sigma}\right)\right), \\
&\qquad \text{for } 0 \leq m \leq N-2 \\
\alpha_{N-1} &= \frac{1}{2}\left(Q\left(\frac{-1-2h_N}{\sigma}\right)+Q\left(\frac{-1+2h_N}{\sigma}\right)\right) \\
\alpha_N &= Q\left(-\frac{1}{\sigma}\right),
\end{aligned}$$

where

$$\beta_m = 2\sum_{i=m+1}^{N}|h_i|.$$

Using these results in equation (12), we can obtain the lower and upper bounds on the probability of error in the steady state.

As an example, consider the channel whose transfer function is given by

$$H(z^{-1}) = 1 + 0.8z^{-1} + 0.6z^{-2} + 0.4z^{-3}.$$

In this case, the feedback section contains three delay elements, $N = 3$. The transition probabilities are given by

$$\begin{aligned}
\alpha_0 &\geq \frac{1}{2}\left(Q\left(\frac{-1-2h_1+\beta_1}{\sigma}\right)+Q\left(\frac{-1+2h_1+\beta_1}{\sigma}\right)\right) \\
\alpha_0 &\leq \frac{1}{2}\left(Q\left(\frac{-1-2h_1-\beta_1}{\sigma}\right)+Q\left(\frac{-1+2h_1-\beta_1}{\sigma}\right)\right) \\
\alpha_1 &\geq \frac{1}{2}\left(Q\left(\frac{-1-2h_2+\beta_2}{\sigma}\right)+Q\left(\frac{-1+2h_2+\beta_2}{\sigma}\right)\right) \\
\alpha_1 &\leq \frac{1}{2}\left(Q\left(\frac{-1-2h_2-\beta_2}{\sigma}\right)+Q\left(\frac{-1+2h_2-\beta_2}{\sigma}\right)\right) \\
\alpha_2 &= \frac{1}{2}\left(Q\left(\frac{-1-2h_3}{\sigma}\right)+Q\left(\frac{-1+2h_3}{\sigma}\right)\right) \\
\alpha_2 &= Q\left(-\frac{1}{\sigma}\right),
\end{aligned}$$

where $\beta_1 = 2(|h_2| + |h_3|)$ and $\beta_2 = 2|h_3|$. Figure 5, below, shows the lower and the upper bound of the probability of error. Also shown is the probability of error computed from simulation. The "no ISI" lower bound is also shown, and one can see that the lower bound derived here is tighter than the "no ISI" bound.

VI. KALMAN-TYPE DECORRELATION ALGORITHM FOR THE BLIND DFE

A drawback of blind equalizer is the speed of convergence. It can take several hundred to several thousand symbols for the blind equalizer to converge. The speed of convergence of the conventional LMS equalizer was improved by using a weighted sum of the past squared errors [23]. The resulting algorithm, known as the Recursive Least Squares (RLS) or Kalman algorithm, improves the speed of convergence substantially. The penalty is an increase in computational complexity.

The decorrelation algorithm described in this paper uses a simple error function, which makes it easy to extend to a Kalman-type algorithm. In this section we show how one can improve the convergence of the algorithm by considering a weighted sum of past correlations. In other words, instead of decorrelating instantaneous realizations of the output, we decorrelate the time-average weighted samples of the output, *i.e.*, we force

$$\sum_{j=0}^{k} \lambda^{k-j} A_j \mathbf{A}_j$$

to zero, where $\mathbf{A'}_j = [A_{j-1}, A_{j-2}, \cdots, A_{j-n}]$. Substituting for A_k from equation (8) and setting the weighted correlation time average to zero, we get

$$\sum_{j=0}^{k} \lambda^{k-j} \mathbf{A}_j \left(X_j - \hat{\mathbf{A}}'_j \mathbf{W}_k \right) = 0,$$

where we have introduced the time index k in the weight vector, since a weight update equation will be derived from the above expression. The above equation leads to

$$\mathbf{W}_k = \mathbf{R^{-1}}_k \mathbf{D}_k, \tag{13}$$

where

$$\mathbf{R}_k \triangleq \sum_{j=0}^{k} \lambda^{k-j} \mathbf{A}_j \hat{\mathbf{A}}'_j$$

and

$$\mathbf{D}_k \triangleq \sum_{j=0}^{k} \lambda^{k-j} X_j \mathbf{A}_j.$$

6.1 The Recursive Matrix Inversion

Equation (13) involves the inversion of an $n \times n$ matrix, $\mathbf{R}_k$, and the Kalman formulation involves a recursion formula for the evaluation of the inverse matrix. A similar one can be used here.

It is important to note that matrix $\mathbf{R}_k$ can be obtained recursively as

$$\mathbf{R}_k = \lambda \mathbf{R}_{k-1} + \mathbf{A}_k \hat{\mathbf{A}}_k'. \tag{14}$$

It is known that for any $\mathbf{A}$ nonsingular matrix, and $\mathbf{u}$ and $\mathbf{v}$ the following is true:

$$(\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}. \tag{15}$$

Therefore, using equation (15) in equation (14), we get a recursive formula for $\mathbf{R^{-1}}_k$:

$$\mathbf{R^{-1}}_k = \frac{1}{\lambda}\left(\mathbf{R^{-1}}_{k-1} - \frac{\mathbf{R^{-1}}_{k-1}\mathbf{A}_k\hat{\mathbf{A}}_k'\mathbf{R^{-1}}_{k-1}}{\lambda + \hat{\mathbf{A}}_k'\mathbf{R^{-1}}_{k-1}\mathbf{A}_k}\right). \tag{16}$$

Next define

$$\mathbf{P}_k \triangleq \mathbf{R^{-1}}_k, \tag{17}$$

and further define the Kalman vector gain as

$$\mathbf{k}_k = \frac{1}{\lambda + \mu_k}\mathbf{P}_{k-1}\mathbf{A}_k, \tag{18}$$

where the scalar μ_k is given by

$$\mu_k = \hat{\mathbf{A}}_k'\mathbf{R^{-1}}_{k-1}\mathbf{A}_k.$$

Using the above definitions, one can write equation (16) as

$$\mathbf{P}_k = \frac{1}{\lambda}\left(\mathbf{P}_{k-1} - \mathbf{k}_k\hat{\mathbf{A}}_k'\mathbf{P}_{k-1}\right). \tag{19}$$

The vector $\mathbf{D}_k$ can also be obtained recursively as

$$\mathbf{D}_k = \lambda\mathbf{D}_{k-1} + X_k\mathbf{A}_k. \tag{20}$$

Using equations (13) and (17), we can write

$$\mathbf{W}_k = \mathbf{P}_k\mathbf{D}_k.$$

Therefore, using the recursive formulae for $\mathbf{P}_k$ and $\mathbf{D}_k$ from equations (19) and (20) respectively, we get

$$\begin{aligned}\mathbf{W}_k &= \frac{1}{\lambda}\left(\mathbf{P}_{k-1} - \mathbf{k}_k\hat{\mathbf{A}}_k'\mathbf{P}_{k-1}\right)(\lambda\mathbf{D}_{k-1} + X_k\mathbf{A}_k) \\ &= \mathbf{P}_{k-1}\mathbf{D}_{k-1} + \frac{1}{\lambda}X_k\mathbf{P}_{k-1}\mathbf{A}_k\end{aligned}$$

$$
\begin{aligned}
& -\mathbf{k}_k \hat{\mathbf{A}}_k' \mathbf{P}_{k-1} \mathbf{D}_{k-1} - \frac{1}{\lambda} X_k \mathbf{k}_k \hat{\mathbf{A}}_k' \mathbf{P}_{k-1} \mathbf{A}_k \\
= \; & \mathbf{W}_{k-1} + \frac{1}{\lambda} X_k \left(\lambda + \mu_k\right) \mathbf{k}_k \\
& -\mathbf{k}_k \hat{\mathbf{A}}_k' \mathbf{W}_{k-1} - \frac{1}{\lambda} X_k \mu_k \mathbf{k}_k \\
= \; & \mathbf{W}_{k-1} + \mathbf{k}_k \left(X_k - \hat{\mathbf{A}}_k' \mathbf{W}_{k-1}\right) \\
= \; & \mathbf{W}_{k-1} + z_k \mathbf{k}_k, \qquad (21)
\end{aligned}
$$

where z_k is given by

$$z_k = \left(X_k - \hat{\mathbf{A}}_k' \mathbf{W}_{k-1}\right).$$

The order that constitutes the Kalman type decorrelation algorithm is summarized below:

$$\mathbf{W}_k = \mathbf{W}_{k-1} + z_k \mathbf{k}_k,$$

where

$$z_k = \left(X_k - \hat{\mathbf{A}}_k' \mathbf{W}_{k-1}\right).$$

The vector K_k is evaluated by the recursions:

$$\mathbf{k}_k = \frac{1}{\lambda + \mu_k} \mathbf{P}_{k-1} \mathbf{A}_k$$

$$\mathbf{P}_k = \frac{1}{\lambda} \left(\mathbf{P}_{k-1} - \mathbf{k}_k \hat{\mathbf{A}}_k' \mathbf{P}_{k-1}\right),$$

where

$$\mu_k = \hat{\mathbf{A}}_k' \mathbf{P}_{k-1} \mathbf{A}_k.$$

6.2 Simulation Results

A channel whose transfer function given by

$$H(z^{-1}) = 1 + 0.5z^{-1} - 1.44z^{-2}$$

was used to demonstrate the improvement in the convergence rate when using the Kalman approach. Figure 6 depicts the learning curve of the decorrelation DFE and the Kalman-type decorrelation DFE, obtained by ensemble-averaging over 100 independent simulation runs. Figure 7 depicts the averaged squared error of the decorrelating DFE and the Kalman-type decorrelating DFE. The increased speed of the latter is clearly shown.

VII. CONCLUSIONS

In this paper we introduced a new criterion and algorithm for blind equalization. This algorithm is used in conjunction with a decision feedback equalizer. It decorrelates the data sequence at the input of the slicer. It was shown to converge to the

optimum weights irrespective of the initial error rate. Examples for minimum and non-minimum phase channels supplement the proven analytical results.

With an adaptation constant μ of 0.01 the algorithm converges after 300 iterations. The simulation also shows that the algorithm converges to the optimum point regardless of the initial setting.

We also used the error model proposed by Duttweiler, *et al.*, to obtain upper and lower bounds on the steady state probability of error for a decision feedback equalizer. The upper bound is essentially the same as that obtained by Duttweiler. The lower bound is tighter than the commonly used "no ISI" lower bound.

We have also derived a Kalman-type decorrelation algorithm for the blind DFE. This has a higher rate of convergence than the original decorrelation algorithm, with the added cost of an increase in the number of computations.

References

[1] Y. Sato: *A Method of Self-Recovering Equalization for Multi-Level Amplitude Modulation.* "IEEE Trans. on Communications", Vol. 23, June 1975, p. 679-682.

[2] D.N. Godard: *Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems.* "IEEE Trans. on Communications", Vol. 28, Nov. 1980, p. 1967-1875.

[3] A. Benveniste, M. Goursat, G. Ruget: *Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications.* "IEEE Trans. on Automatic Control", Vol. 25, June 1980, p. 385-399.

[4] S. Bellini, F. Rocca: *Blind Deconvolution: Polyspectra or Bussgang Techniques.* In: *Digital Communications.* (E. Biglieri and G. Prati, Ed.), North Holland, Elsevier Science Publishers B.V. 1986.

[5] O. Shalvi, E. Weinstein: *New Criteria for Blind Deconvolution of non-minimum Phase Systems (Channels).* "IEEE Trans. on Information Theory", Vol. 36, March 1990, p. 312-321.

[6] J.R. Treichler, M.G. Agee: *A New Approach to Multipath Correction of Constant Modulus Signals.* "IEEE Trans. on Acoustics, Speech, and Signal Processing", Vol. 31, pp. 349-472, April 1983.

[7] J.R. Treichler, M.G. Larimore: *New Processing Techniques Based on the Constant Modulus Adaptive Algorithm.* "IEEE Trans. on Acoustics, Speech, and Signal Processing", Vol. 33, April 1985, p. 420-431.

[8] C.R. Johnson: *Admissibility in Blind Adaptive Channel Equalization.* "IEEE Control Systems Magazine", Jan. 1991, p. 3-15.

[9] Z. Ding, R.A. Kennedy, B.O. Anderson, C.R. Johnson: *Ill-Convergence of Godard Blind Equalizers in Data Communications.* "IEEE Trans. on Communications", Vol. 39, Sept. 1991, p. 1313-1328.

[10] Z. Ding, R.A. Kennedy, B.O. Anderson, C.R. Johnson: *Local Convergence of the Sato Blind Equalizer and Generalizations Under Practical Constraints.* "IEEE Trans. on Information Theory", Vol. 129, Jan. 1993, p. 129-144.

[11] Z. Ding: *Application Aspects of Blind Adaptive Equalizers in QAM Data Communications.* Ph.D. Dissertation, Cornell University, 1990.

[12] S. Verdu, B. Anderson, R.A. Kennedy: *Anchored Blind Equalization. 25th Conf. Information Sciences and Systems.* Baltimore (USA). March 1991, p. 774-779.

[13] S. Vembu, S. Verdu, R.A. Kennedy, W.A. Sethares: *Convex Cost Functions in Blind Equalization. 25th Conf. Info. Sci. and Systems.* Baltimore (USA). March 1991, p. 792-797.

[14] R.A. Kennedy Z. Ding: *Blind Adaptive Equalizers for Quadrature Amplitude Modulated Communication Systems Based on Convex Cost Functions.* "Optical Engineering", Vol. 31, June 1992, p. 1189-1199.

[15] B. Porat, B. Friedlander: *Blind Equalization of Digital Communication Channels Using High-Order Moments.* "IEEE Trans. on Signal Processing", Vol. 39, Feb. 1991, p. 522-526.

[16] D. Hatzinakos, C. Nikias: *Blind Equalization Using a Tricepstrum-Based Algorithm.* "IEEE Trans. on Communications", Vol. 39, May 1991, p. 669-682.

[17] N. Seshadri: *Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques. GLOBECOM '90.* San Diego (USA). Dec. 1990, p. 807.1.1-807.1.5.

[18] M. Ghosh, C. Weber: *Maximum-Likelihood Blind Equalization.* "Optical Engineering", Vol. 31, June 1992, p. 1224-1228.

[19] M. Feder, J.A. Capitovic: *Algorithms for Joint Channel Estimation and Data Recovery-Application to Equalization in Underwater Communications.* "IEEE Journal of Ocean Engineering", Vol. 16, Jan. 1991, p. 42-55.

[20] R.A. Kennedy, G. Pulford, B.D.O. Anderson, R. R. Bitmead: *When Has a Decision-Directed Equalizer Converged?* "IEEE Trans. on Communications", Vol. 37, August 1989, p. 879-884.

[21] D.L. Duttweiler, J.E. Mazo, D.G. Messerschmidt: *An Upper Bound on the Error Probability in Decision-Feedback Equalization.* "IEEE Trans. on Information Theory", Vol. 20, July 1974, p. 490-497.

[22] R.E. Kamel, Y. Bar-Ness: *Error Performance of the Blind Decision Feedback Equalizer Using Decorrelation. Milcom '94.* Ft. Monmouth (USA). Oct. 1994, p. 618-622.

[23] D. Godard: *Channel Equalization Using a Kalman Filter for Fast Data Transmission.* "IBM Journal of Research and Development", Vol. 18, May 1974, p. 267-273.

[24] R.D. Gitlin, F.R. Magee: *Self-Orthogonalizing Adaptive Equalization Algorithms.* "IEEE Trans. on Communications", Vol. 25, July 1977, p. 666-672.

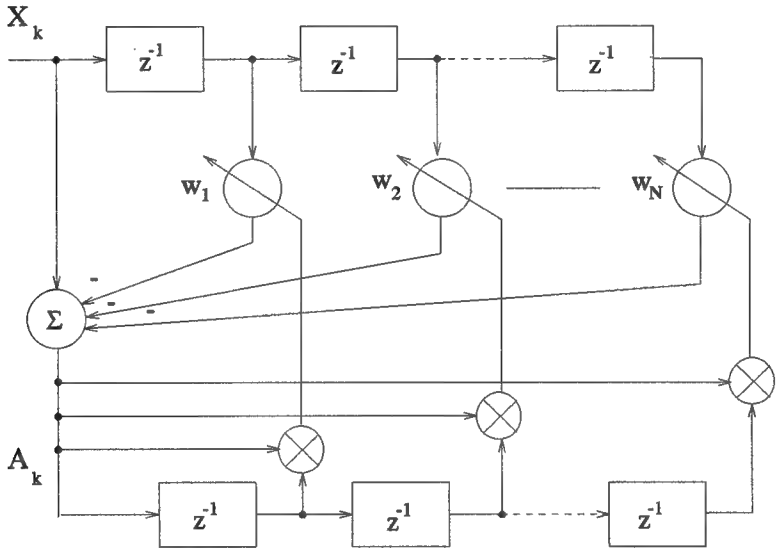


Figure 1: Blind Linear Equalizer

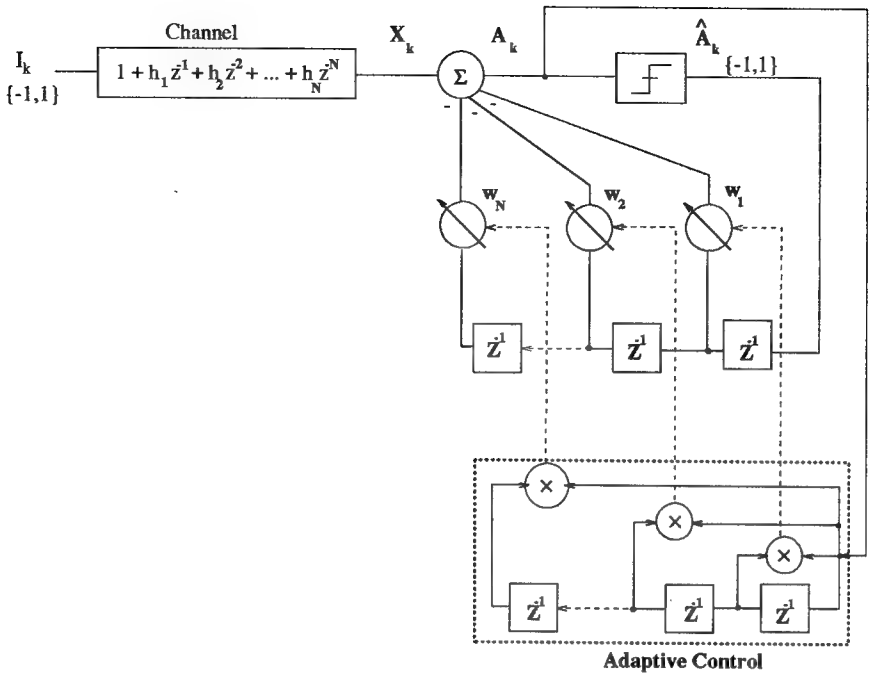


Figure 2: Decision Feedback Equalizer with Decorrelation Control

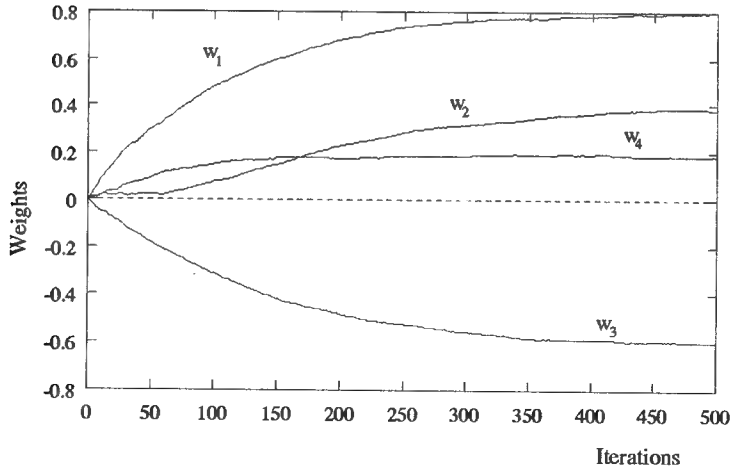


Figure 3: Learning Curve of the BDFE

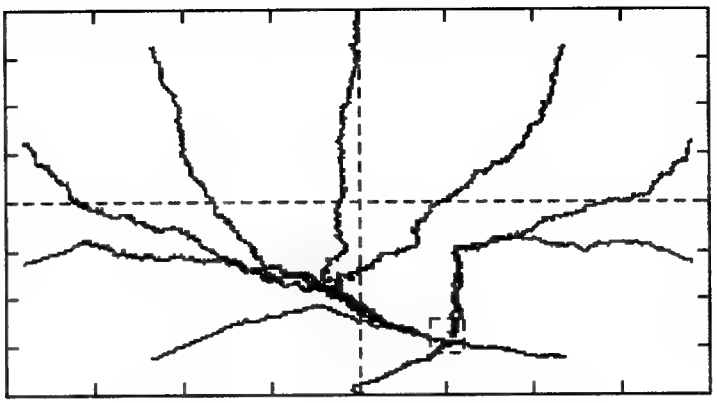


Figure 4: Convergence of the BDFE

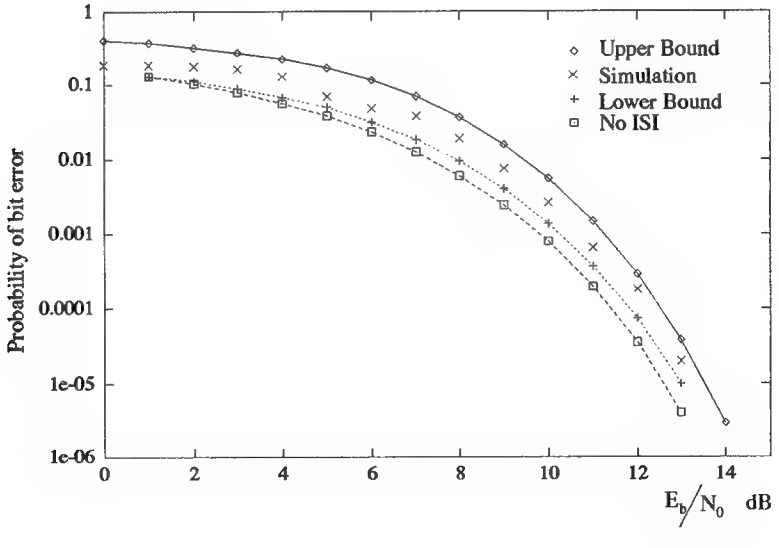


Figure 5: Probability of Error

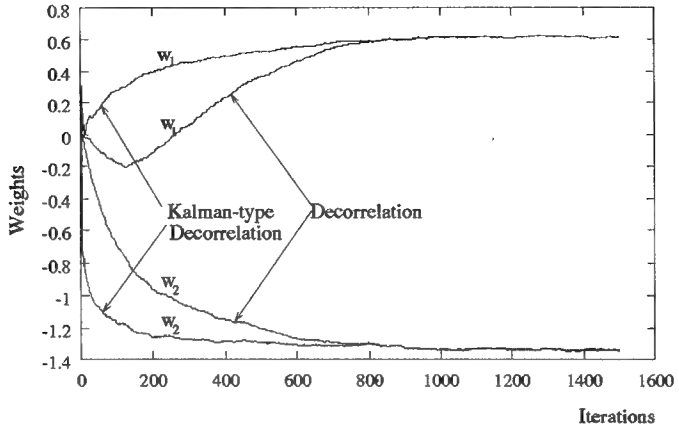


Figure 6: Learning Curves of the Decorrelation DFE and Kalman-type Decorrelation DFE

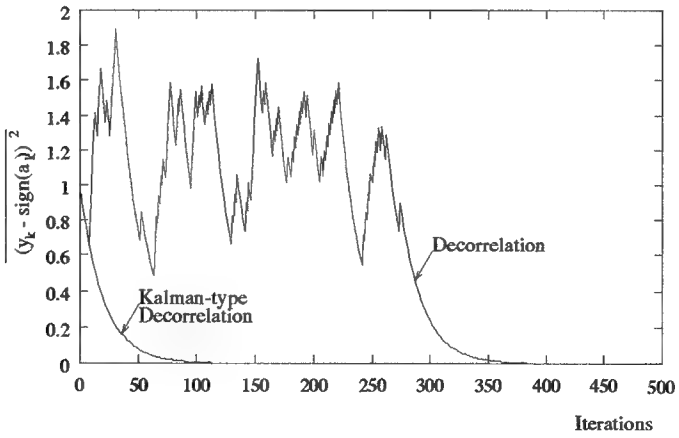


Figure 7: Residual ISI of the Decorrelating DFE and Kalman-type Decorrelation DFE

APPENDIX C: PART II

ANCHORED BLIND EQUALIZATION USING THE CONSTANT MODULUS ALGORITHM

by
Raafat Kamel and Yeheskel Bar-Ness

ABSTRACT

Blind equalization is a technique of adapting an equalizer without the need of a training sequence. The constant modulus algorithm (CMA) is one of the first known blind equalization algorithms. The cost function of the CMA exhibits local minima, which are the primary cause of the ill-convergence of the CMA. Anchoring the CMA improves the performance of the CMA in terms of ill- convergence. This technique is used with the linear and the decision feedback equalizers. It is shown that the adaptive equalizer will always remove intersymbol interference (ISI) as long as the channel gain exceeds a certain critical value.

I. INTRODUCTION

Adaptive blind channel equalizers eliminate the use of training signals in digital communication when their transmission is impractical. The problem of adaptive blind equalization is that of finding an appropriate cost function (or equivalent error function) that reflects the amount of intersymbol intersymbol (ISI) introduced by the channel, and which does not involve the transmitted symbols [1-5]. Optimization of the cost function should lead to minimization of the ISI. Blind equalization algorithms differ in the choice of the cost function.

The first known blind equalization algorithm was introduced by Sato [1]. Sato's blind equalization algorithm error function was later generalized by Benveniste, *et al.,* [2]. Considering Sato's cost function, Godard [3] then described a class of cost functions that generalizes Sato's. A similar scheme, known as the constant modulus algorithm (CMA) was developed

separately by Treichler, *et al.*, [6-7] based on property restoral concept.

Bellini, *et al.* [4], followed a different approach and developed what they termed "Bussgang Techniques." Based on some assumptions about the equalizer and the channel parameters, they derived a maximum likelihood estimator of the reference signal. This estimator depends on the type of modulation used and the signal-to-noise ratio (SNR). The aforementioned algorithms are viewed as special cases of the Bussgang technique.

The above algorithms used non-convex cost functions which possess local minima to which the blind equalizer might converge. Some of these equilibria may be undesirable, *i.e.*, at those equilibria the equalizer will not be able to remove ISI. This was demonstrated by Ding, *et al.*, for the Godard algorithm [9] and for the Sato algorithm [8]. In [10], the ill convergence of the Benveniste, *et al.*, algorithms [2] was also considered, thus proving that none of the previous algorithms were globally convergent. For these algorithms equalizer initialization becomes an important issue. One would initialize the equalizer away from the neighborhood of the ill-convergent minima.

Verdu, *et al.*, [11] developed a technique that insures global convergence of blind equalizers. The key observation in [11] is that over-parameterizing the blind equalizer is the prime cause of ill-convergence, and hence anchoring (setting the first coefficient to one) the blind equalizer was proposed [11-12]. This, together with using a convex function, guarantees convergence. Verdu used the minimized energy as a cost function. Vembu, *et al.* [12], used the l_1 norm of the equalizer weight as a cost function, which was approximated by the l_p norm of the equalizer output. Kennedy and Ding [13] applied the concept of anchored equalization to a QAM transmission. The cost function used was the l_p norm of the real or imaginary part of the equalizer's output.

Another family of blind equalization algorithms that appears in the literature is based on high-order moments and polyspectra [14] and [15]. In general, these algorithms give better performance at the expense of higher arithmetic complexity. Basically, they use the received samples to estimate the channel parameters and reconstruct the transmitted data via inverse filtering. The computational complexity of these algorithms makes them inappropriate for on-line processing.

Maximum likelihood blind sequence estimation has also been proposed [16-19]. These approaches involves the use of a blind channel identification scheme is conjunction with the Viterbi Algorithm.

A new approach to blind channel identification was reported in [20-21], that exploits the cyclostationery nature of digital signals. Interested readers are refered to [20-21].

In this paper we will consider using the CMA with an anchored equalizer. This approach will improve the performance of the convergence property of the original CMA. It is shown that the anchored blind equalizer with the CMA (ACMA) converges to the channel parameters rendering zero ISI provided the channel gain exceeds a certain critical value. If the gain drops below this critical point, the algorithm will converge to a local minimum. This problem can be alleviated by introducing a gain in the equalizer. The speed of convergence of this equalizer will be compared to that of AMEA [11].

This paper is organized as follows. First we consider using the ACMA for blind equalization of autoregressive and moving average type channels in sections 2 and 3, respectively. In section 4, we present simulation results. Conclusions are given in section 5.

II. EQUALIZATION OF AUTOREGRESSIVE CHANNELS

Consider an AR channel of order n ($AR(n)$), the received signal is given by

$$r_k = g a_k + \sum_{i=1}^{n} \alpha_i r_{k-i},$$

where g is an arbitrary gain, and α_i's are the $AR(n)$ channel parameters. The information symbols (a_k's) are binary, independent and identically distributed with zero mean and unit variance. The moving average (MA) anchored equalizer output has its first tap set to 1, and, therefore, its output is given by

$$\begin{aligned} y_k &= r_k + \sum_{i=1}^{n} w_i r_{k-i} \\ &= g a_k + \sum_{i=1}^{n} (\alpha_i + w_i) r_{k-i} \\ &\triangleq g a_k + isi_k \end{aligned} \tag{1}$$

where w_i's are the equalizer's coefficients and isi_k is given by

$$isi_k = \sum_{i=1}^{n} (\alpha_i + w_i) r_{k-i}. \tag{2}$$

The CMA exploits the fact that the original constellation has a constant envelope, that is, $E\{|a_k|\} = 1$ for all k. Therefore an appropriate cost function would be

$$J(y_k) = \left(|y_k|^2 - 1\right)^2, \tag{3}$$

which is minimized when the equalizer output has a constant modulus ($E\{|y_k|\} = 1$). Using stochastic gradient descent to minimize the above cost function, the update equation for the CMA is given by

$$w_i^{k+1} = w_i^k - \mu r_{k-i} \left(y_k^2 - 1\right) y_k \quad \text{for } i = 1, \cdots, n. \tag{4}$$

Figure 1 shows the anchored linear equalizer and the CMA control.

Multiply the above equation by $\alpha_i + w_i^k$:

$$w_i^{k+1}\left(\alpha_i + w_i^k\right) = w_i^k\left(\alpha_i + w_i^k\right) - \mu\left(\alpha_i + w_i^k\right) r_{k-i} y_k \left(y_k^2 - 1\right).$$

Now take the expectation of the above conditioned on w_i^k:

$$\alpha_i E\{w_i^{k+1}|w_i^k\} + w_i^k E\{w_i^{k+1}|w_i^k\} = \alpha_i w_i^k + \left(w_i^k\right)^2 - \mu\left(\alpha_i + w_i^k\right) E\{r_{k-i} y_k \left(y_k^2 - 1\right)|w_i^k\}.$$

Steady state is reached when $E\{w_i^{k+1}|w_i^k\} = w_i^k$, and, therefore, we have

$$\left(\alpha_i + w_i^k\right) E\{r_{k-i} y_k \left(y_k^2 - 1\right)|w_i^k\} = 0 \quad \text{for} \quad i = 1, \cdots, n \text{ and all } k. \tag{5}$$

The above n equations determine the points of equilibrium of the algorithm. One would have to solve the above equations in order to determine whether the algorithm would converge to the desired values ($w_i = \alpha_i$) and, hence, cancel the ISI completely. Instead we show directly that, under certain conditions for the gain g, equation (5) implies complete cancellation of ISI.

Adding the above n equations we get

$$\sum_{i=1}^{n}\left(\alpha_i + w_i^k\right) E\{r_{k-i} y_k \left(y_k^2 - 1\right)|w_i^k\} = 0.$$

Substituting equations (1) and (2) in the above equation results in

$$\begin{aligned} & E\{isi_k\left((ga_k + isi_k)^2 - 1\right)(ga_k + isi_k)|w_i^k\} \\ = & E\{isi_k\left(g^3 a_k^3 + \left(3g^2 a_k^2 - 1\right) isi_k + \left(3isi_k^2 - 1\right) ga_k + isi_k^3\right)|w_i^k\}\} \\ = & 0. \end{aligned}$$

Now, with the definition (2), isi_k depends on the previous data and w_i^k (which itself depends only on the previous data $a_{k-i}, i \geq 1$), therefore isi_k is independent of the current data a_k. Using this together with the fact that both have a zero mean, we get

$$E\left\{\left(y_k^2 - 1\right) y_k \sum_{i=1}^{n}\left(\alpha_i + w_i\right) r_{k-i}|w_i^k\right\} = E\left\{isis_k(y_k^2 - 1)y_k|w_1^k\right\}$$

$$
\begin{aligned}
&= \left(3g^2\sigma_a^2 - 1\right) E\{isi_k^2|w_i^k\} + E\{isi_k^4|w_i^k\} \\
&= \left(3g^2 - 1\right) E\{isi_k^2|w_i^k\} + E\{isi_k^4|w_i^k\} \qquad (6) \\
&= 0.
\end{aligned}
$$

If $3g^2 - 1$ in equation (6) is a positive quantity then it can be written as

$$E\{isi_k^4|w_i^k\} = -K^2 E\{isi_k^2|w_i^k\} \qquad (7)$$

with K positive. However, both $E\{isi_k^4|w_i^k\}$ and $E\{isi_k^2|w_i^k\}$ being positive quantities implies

$$E\{isi_k^4|w_i^k\} = E\{isi_k^2|w_i^k\} = 0,$$

and together with the fact that the expected value of isi_k is 0, we conclude that $isi_k = 0$ with a probability of 1. In sum, if the algorithm reaches a steady state then equation (5) is satisfied for $i = 1, \cdots, n$, and from equation (1)

$$r_k = ga_k \qquad \text{for all } k.$$

If, however, $3g^2 - 1$ is negative then equation (6) can be written as

$$E\{isi_k^4|w_i^k\} = \Lambda^2 E\{isi_k^2|w_i^k\}, \qquad (8)$$

where $\Lambda^2 = 1 - 3g^2$, in which case the ISI power is not necessarily zero. This corresponds to a case wherein the algorithm converges to a local minimum, which could be undesirable. We show the existence of such undesirable equilibria by using a simple example.

2.1 Undesirable Equilibria

Consider an $AR(n)$ channel with one feedback tap, given by

$$r_k = ga_k + \alpha r_{k-n}.$$

The equalizer output is then given by

$$y_k = g a_k + (w + \alpha)\, r_{k-n}, \tag{9}$$

and the ISI term by,

$$isi_k = (w + \alpha) r_{k-n}. \tag{10}$$

It is then easy to show that

$$\frac{E\{r_k^4\}}{E\{r_k^2\}} = g^2 \frac{1 + 5\alpha^2}{1 - \alpha^4},$$

and by substituting equation (10),

$$\frac{E\{isi_k^4 | w\}}{E\{isi_k^2 | w\}} = g^2 (w + \alpha)^2 \frac{1 + 5\alpha^2}{1 - \alpha^4}. \tag{11}$$

Combining equation (11) with equation (8), we get

$$\begin{aligned} (w + \alpha)^2 &= \Lambda^2 \frac{1 - \alpha^4}{g^2(1 + 5\alpha^2)} \\ &= \left(1 - 3g^2\right) \frac{1 - \alpha^4}{g^2(1 + 5\alpha^2)}, \end{aligned} \tag{12}$$

or

$$w = -\alpha \pm \frac{\sqrt{1 - 3g^2}}{g} \sqrt{\frac{1 - \alpha^4}{(1 + 5a^2)}}. \tag{13}$$

This clearly shows that the weight w will not converge to the correct channel parameter α.

In particular, following a procedure similar to that of [9], we predict the condition on g and α which results in $w = 0$, and in no ISI cancellation (see equation (9)). Setting $w = 0$, in equation (13), we get

$$g^2 = \frac{1 - \alpha^4}{3 + \alpha^2 + 2\alpha^4}.$$

Therefore if the gain satisfies the above equation, the equalizer will not remove ISI. In conclusion, if the condition $3g^2 - 1 > 0$ is guaranteed, one would ensure that the algorithm would always converge to the correct channel parameters. In other words should the channel gain g be less than $\frac{1}{\sqrt{3}}$, the algorithm will not converge to the no ISI case.

The actual dependence of steady state and ill convergence of the ACMA on the channel gain g is examined in the following example. For the $AR(1)$

$$r_k = ga_k + 0.6r_{k-1}. \tag{14}$$

The ACMA equalizer is given by

$$y_k = r_k + w^{(k)}r_{k-1}. \tag{15}$$

Using these two equations, we plot in Figure 2 the cost function $J(y_k) = E\{(y_k^2 - 1)^2\}$ as a function of w and g. From this figure it is clear that if the gain $g > \frac{1}{\sqrt{3}} = 0.577$ then the cost function has a unique minimum at $w = \alpha = 0.6$. If, however, the gain g drops below $\frac{1}{\sqrt{3}}$, the function will have two minima and a maximum at $w = \alpha$ and therefore the equalizer will not converge to the channel parameter.

One can alleviate this problem by introducing an arbitrary gain, G, in the equalizer. The output of the equalizer is then given by

$$\begin{aligned} y_k &= Gr_k + G\sum_{i=1}^{n} w_i r_{k-i} \\ &= Gga_k + G\sum_{i=1}^{n}(\alpha_i + w_i). \end{aligned} \tag{16}$$

Following a similar procedure as the one above, one can show that at the algorithm's steady state, ISI is cancelled (with a probability of 1) if and only if

$$g > \frac{1}{G\sqrt{3}}. \tag{17}$$

Thus, one can choose G appropriately such that condition (17) is satisfied. In other words, one would choose G such that for the worst case g condition (17) is met. If the worst case g is 0.01, for example, one would choose $G > \frac{100}{\sqrt{3}}$.

III. EQUALIZATION OF MOVING AVERAGE TYPE CHANNELS

Consider a real MA type of channel of order n, $MA(n)$, the received signal, is given by

$$r_k = ga_k + \sum_{i=1}^{n} h_i a_{k-i},$$

where g is an arbitrary gain of the first cursor and h_i's are the MA(n) channel parameters. The input to the slicer of the decision feedback equalizer is given by

$$\begin{aligned} y_k &= r_k - \sum_{i=1}^{n} w_i \hat{y}_{k-i} \\ &= ga_k + \sum_{i=1}^{n} h_i a_{k-i} - \sum_{i=1}^{n} w_i \hat{y}_{k-i}, \end{aligned} \tag{18}$$

where w_i's are the equalizer's coefficients. Figure 3 shows the decision feedback equalization, with the control section.

Now, if we denote the set of all correct decisions by Y_1 and the set of all incorrect decisions by Y'', *i.e.*,

$$Y' = \{\hat{y}_i : \hat{y}_i = a_i\}$$

$$Y'' = \{\hat{y}_i : \hat{y}_i = -a_i\},$$

then equation (18) can be written as

$$\begin{aligned} y_k &= ga_k + \sum_{i:\hat{y}_{k-i}\epsilon Y'} (h_i - w_i^k)\hat{y}_{k-i} - \sum_{i:\hat{y}_{k-i}\epsilon Y''} (h_i + w_i^k)\hat{y}_{k-i} \\ &= ga_k + \sum_{i=1}^{n} \gamma_i \hat{y}_{k-i} \\ &\triangleq ga_k + isi_k, \end{aligned} \tag{19}$$

where

$$isi_k = \sum_{i=1}^{n} \gamma_i \hat{y}_{k-i},$$

and the γ_i's are given by

$$\begin{aligned} \gamma_i &= (h_i - w_i^k) \qquad \text{for } i : \hat{y}_{k-i}\epsilon Y_1 \\ \gamma_i &= -(h_i + w_i^k) \qquad \text{for } i : \hat{y}_{k-i}\epsilon Y_2. \end{aligned}$$

Applying the constant modulus cost function in equation (3) for the DFE, and using equation (18) for dy_k/dw_i, we get the following update equation

$$w_i^{k+1} = w_i^k + \mu \hat{y}_{k-i} \left(y_k^2 - 1\right) y_k \text{ for } i = 1, 2, \cdots, n. \tag{20}$$

Multiplying equation (20) by γ_i,

$$w_i^{k+1}\gamma_i = w_i^k\gamma_i + \mu\gamma_i \hat{y}_{k-i} \left(y_k^2 - 1\right) y_k. \tag{21}$$

As before, take expectation conditioned on w_i^k to obtain

$$E\{w_i^{k+1}\gamma_i | w_i^k\} = w_i^k E\{\gamma_i | w_i^k\} + \mu E\{\gamma_i \hat{y}_{k-i} \left(y_k^2 - 1\right) y_k | w_i^k\}. \tag{22}$$

It is simple to show that steady state is reached (that is $E\{w_i^{k+1}|w_i^k\} = w_i^k$) if and only if

$$E\{\gamma_i \hat{y}_{k-i} \left(y_k^2 - 1\right) y_k\} = 0 \text{ for all } i \text{ for which } \hat{y}_{k-i} \epsilon Y_1 \text{ or } \hat{y}_{k-i} \epsilon Y_2.$$

Summing on i we have

$$\begin{aligned} E\left\{\sum_{i=1}^{n} \gamma_i \hat{y}_{k-i} \left(y_k^2 - 1\right) y_k | w_i^k\right\} &= E\{isi_k \left(y_k^2 - 1\right) y_k | w_i^k\} \\ &= 0. \end{aligned}$$

This is exactly the same equation we have for the AR(n) channel case. In Figure 4.3 we present a digital implementation of the DFE-ACMA for an MA type channel.

IV. SIMULATION

In this section we present simulation results of the anchored constant modulus algorithm as it is applied to the linear equalization for autoregressive channels and to the decision feedback equalization of moving average type channels. In particular, we show the effect of the gain g on the performance.

4.1 Linear Equalization

Consider the $AR(1)$ channel whose output signal r_k is given by

$$r_k = ga_k + 0.6r_{k-1}$$

The linear equalizer taps weights are updates using

$$w^{(k+1)} = w^{(k)} - \mu r_{k-1} y_k \left(y_k^2 - 1\right).$$

The averaged squared error of this equalizer is given in Figure 4

4.2 Decision Feedback Equalization

To examine the performance of the DFE we consider a channel whose transfer function is given by

$$H(z^{-1}) = g + 0.5z^{-1} - 1.44z^{-2} \tag{23}$$

with the corresponding adaptation rule of equation (20). The averaged squared error is depicted in Figure 5.

In these figures, the estimate of the residual ISI power is obtained by passing the sequence of the squared error$((a_k - \text{ sgn } (y_k))^2)$ through a smoothing filter whose transfer function is given by $0.05/\left(1 - 0.95z^{-1}\right)$. These figures show clearly that the speed of convergence of the ACMA, for the linear and decision feedback equalizers, depends on the channel gain g. As the gain g the approaches breakpoint, $g = \frac{1}{\sqrt{3}}$, the algorithm takes a longer time to converge. We notice that for $1 \leq g \leq 0.7$, the speed of convergence is nearly constant, reaching approximately zero after 130 iterations, while for $g = 0.6$ the algorithm converges after 250 iterations. The ill convergence of the algorithm is also evident for the gain of $g = 0.5 < \frac{1}{\sqrt{3}}$.

The performance of the ACMA is also compared with the anchored minimum energy algorithm described in [11]. In Figure 6, we depict the convergence of this algorithm when used with the $AR(1)$ channel used in equation (9). The adaptation rule for the AMEA is given by

$$w^{(k+1)} = w^{(k)} + \mu y_k r_{k-1}$$

As in ACMA, the speed of convergence depends on the channel gain g. However, being globally convergent, the AMEA shows no ill convergence for a small g.

Finally, in Figure 7 we compare the rate of convergence of the ACMA with that of the AMEA for $g = 1.0$. The ACMA converges faster than the AMEA.

V. CONCLUSION

In this paper we used the concept of anchoring the blind equalizer with the constant modulus algorithm for AR and MA channels. We showed analytically and through simulation that the algorithm converges successfully if the unknown channel gain exceeds a certain value ($\frac{1}{\sqrt{3}}$). The algorithm will fail to converge to the desired value if the channel gain drops below this value. This problem can be minimized if we introduce a gain in front of the equalizer. Introducing a gain at the equalizer will not eliminate the problem, but it will lower the critical point below which ill convergence appears.

References

[1] Y. Sato, "A Method of Self-Recovering Equalization for Multi-Level Amplitude Modulation," *IEEE Trans. on Communications*, COM-23, pp. 679-682, June 1975.

[2] A. Benveniste, M. Goursat, and G. Ruget, "Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications," *IEEE Trans. on Automatic Control*, AC-25, pp. 385-399, June 1980.

[3] D.N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. on Communications*, COM-28, pp. 1967-1875, November 1980.

[4] S. Bellini and F. Rocca, "Blind Deconvolution: Polyspectra or Bussgang Techniques," In E. Biglieri and G. Prati, editors, *Digital Communications,* pp. 251-263, North-Holland, 1986, Elsevier Science Publishers B. V.

[5] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of non-minimum Phase Systems (Channels)," *IEEE Trans. on Information Theory*, Vol. 36, March 1990, p. 312-321.

[6] J.R. Treichler and M.G. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 31, pp. 349-472, April 1983.

[7] J.R. Treichler and M.G. Larimore, "New Processing Techniques Based on the Constant Modulus Adaptive Algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 33, April 1985, p. 420-431.

[8] Z. Ding, R.A. Kennedy, B.O. Anderson, and C.R. Johnson, "Local Convergence of the Sato Blind Equalizer and Generalizations Under Practical Constraints," *IEEE Trans. on Information Theory*, IT-129 pp. 129-144, Jan. 1993.

[9] Z. Ding, R.A. Kennedy, B.O. Anderson, and C.R. Johnson, "Ill-Convergence of Godard Blind Equalizers in Data Communications," *IEEE Trans. on Communications*, COM-39, pp. 1313-1328, Sept. 1991.

[10] Z. Ding, *Application Aspects of Blind Adaptive Equalizers in QAM Data Communications*, Ph. D. Dissertation, Cornell University, 1990.

[11] S. Verdu, B. Anderson, and R.A. Kennedy, "Anchored Blind Equalization," *1991 Conf. Information Sciences and Systems,* Baltimore, MD. Mar. 1991, pp. 774-779.

[12] S. Vembu, S. Verdu, R.A. Kennedy, and W.A. Sethares, " Convex Cost Functions in Blind Equalization," *Proc. 25th Conf. Info. Sci. and Systems,* pp. 792-797, Baltimore, MD, March 1991.

[13] R.A. Kennedy and Z. Ding, "Blind Adaptive Equalizers for Quadrature Amplitude Modulated Communication Systems Based on Convex Cost Functions," *Optical Engineering,* Vol. 31, pp. 1189-1199, June 1992.

[14] B. Porat and B. Friedlander, "Blind Equalization of Digital Communication Channels Using High-Order Moments," *IEEE Trans. on Signal Processing*, SP-39, pp. 522-526, Feb. 1991.

[15] D. Hatzinakos and C. Nikias, "Blind Equalization Using a Tricepstrum-Based Algorithm," *IEEE Trans. on Communications*, COM-39, pp. 669-682, May 1991.

[16] N. Seshadri, "Joint Data and Channel Estimation Using Fast Blind Trellis Search Techniques," *GLOBECOM '90*, pp. 807.1.1-807.1.5

[17] M. Ghosh and C. Weber, "Maximum-likelihood blind equalization, *Optical Engineering* vol. 31, pp. 1224-1228.

[18] M. Feder and J.A. Capitovic, "Algorithms for Joint Channel Estimation and Data Recovery-Application to Equalization in Underwater Communications," *IEEE Journal of Ocean Engineering,* Vol. 16, pp. 42-55, Jan. 1991.

[19] R.E. Kamel and Y. Bar-Ness, "Blind Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *Electronics Letters*, to appear.

[20] Zhi Ding, "Blind Channel Identification and Equalization Using Spectral Correlation Measurements, Part I: Frequency-Domain Analysis," Article 4 in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed., pp. 417-436, IEEE Press, New York, NY (1994).

[21] Lang Tong, Guanghan Xu and Thomas Kailath, "Blind Channel Identification and Equalization Using Spectral Correlation Measurements, Part II: a Time-Domain Approach," Article 5 in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed., pp. 417-436, IEEE Press, New York, NY (1994).

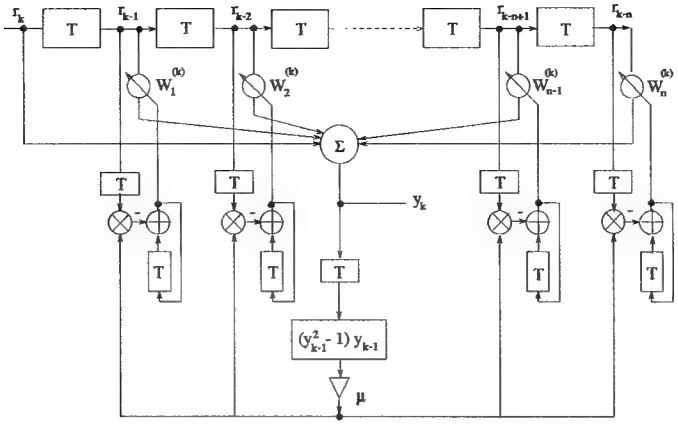


Figure 1: Anchored Linear Equalizer with the CMA

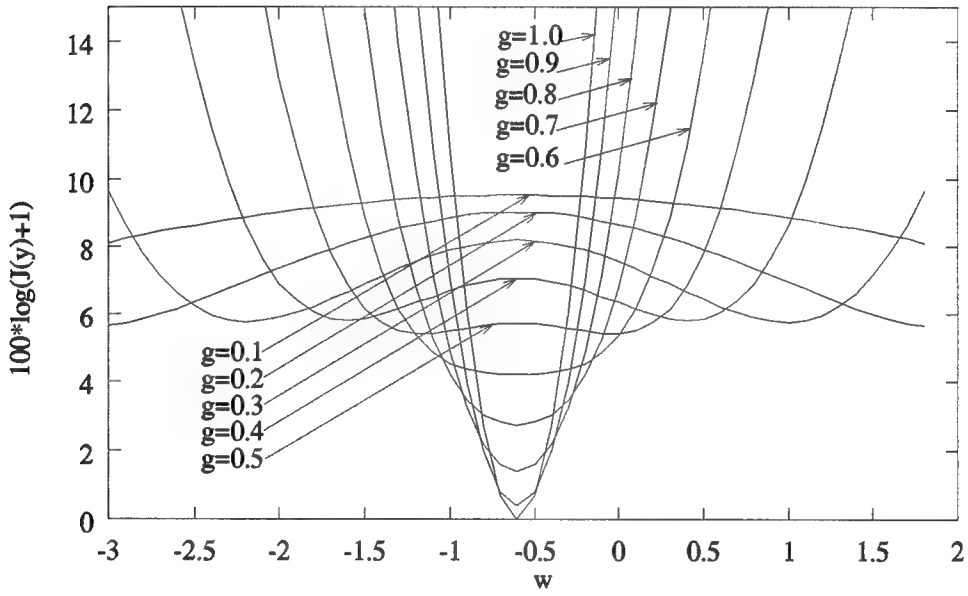


Figure 2: Cost Function for Different Gain g

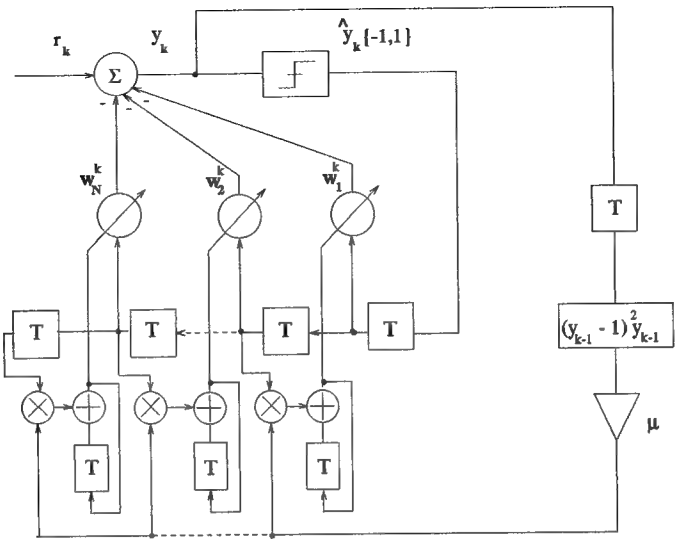


Figure 3: The Anchored DFE using the CMA

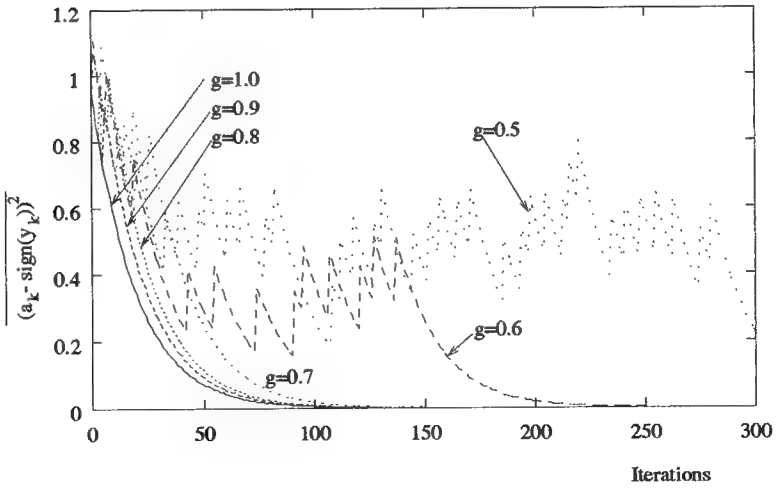


Figure 4: Mean of the Squared Error of ACMA for Different Gain g.

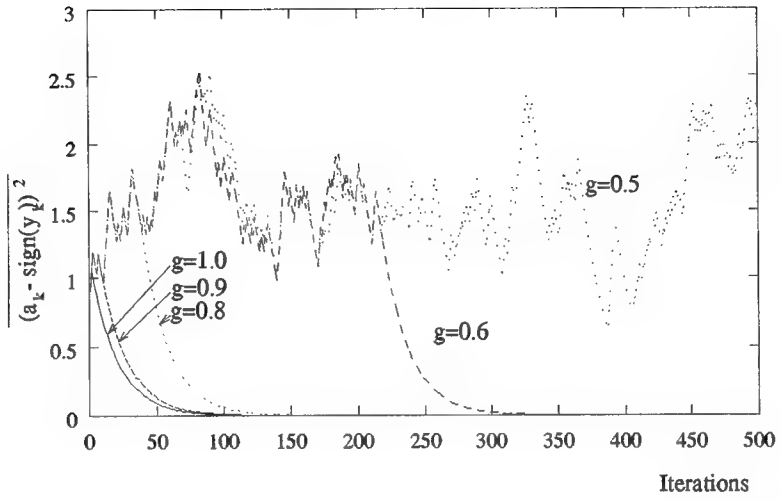


Figure 5: Mean of the Squared Error of ACMA for Different Gain g.

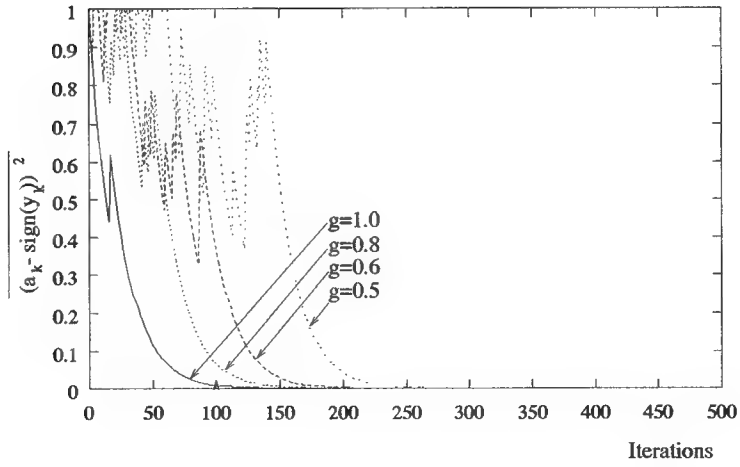


Figure 6: Mean of the Squared Error of AMEA for Different Gain g.

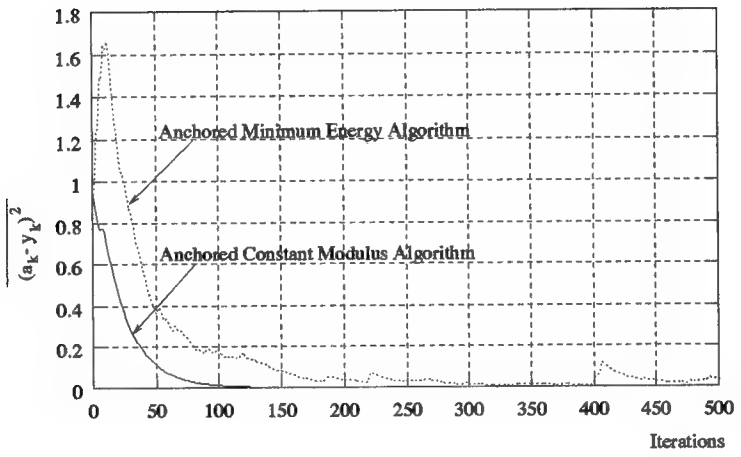


Figure 7: Mean of Error Square of the ACMA and AMEA

APPENDIX C: PART III

BLIND MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION OF DIGITAL SEQUENCES IN THE PRESENCE OF INTERSYMBOL INTERFERENCE

by

Raafat Kamel and Yeheskel Bar-Ness

ABSTRACT

A blind maximum likelihood (ML) sequence estimator for an unknown linear dispersive channels is described. The estimator assumes a channel model with quantized parameters. Two trellises, a channel trellis and a data trellis are defined, to search for the ML channel and data estimates using the Viterbi Algorithm (VA). This approach provides a good performance versus complexity tradeoff.

I. INTRODUCTION:

Adaptive blind equalizers are used to combat intersymbol interference in situations when it is not appropriate to use training sequence in the adaptation of conventional equalizers [2]. Most of the research in blind equalization is geared towards finding suitable cost function for the linear equalizer. For severely distorted channels, linear equalizers enhances noise, resulting in unsatisfactory performance. For these channels blind decision feedback equalization and blind maximum likelihood sequence estimation (MLSE) are prefered. In this communication we consider the latter technique. It is well known that MLSE outperforms both the decision feedback and linear equalizers. Usually the Viterbi algorithm (VA) is used to compute the MLSE efficiently [1], but not without large computation complexity.

For the problem in hand joint ML estimation of the unknown channel and the transmitted data sequence is needed. Seshadri in [3] calculated the least square estimate of channel for every possible sequence. The data sequence that achieves the lowest least square error is then chosen. The resulting complexity in this approach was considerably higher than that of the conventional VA. Ghosh and Weber in [4] developed an iterative algorithm that estimates the channel parameters. An initial guess of the unknown channel is made and the VA is used to estimate the ML of a fixed frame of data. This frame is then used to compute the least square fit of the channel. The process is iterated till the channel estimate converges, the the algorithm is switched to the conventional mode of operation. The frame length needed was long, which incerases the complexity.

To facilitate the proposed approach of this letter, we assume a quantized channel[1] and develop two trellises one for the channel and the other for the data. The VA is used to search the two trellis for the ML channel and data sequence estimate. The output of one is fed into the metric calculator of the other. The resulting scheme offers a considerable reduction in the

[1]This is justified in practice, since finite precision processors are used to implement the algorithm.

computation complexity compared to [3,4]. It also prevails good complexity/performance tradeoff.

II. CHANNEL MODEL AND THE PROPOSED TECHNIQUE

The sampled output of the channel, r_k, at instant k is given by

$$r_k = \mathbf{I}_k'\mathbf{h} + n_k, \tag{1}$$

where $\mathbf{I}_k = [I_k, I_{k-1}, \cdots, I_{k-L}]'$ and $\mathbf{h} = [h_0, h_1, \cdots, h_L]'$.

$\{h_i\}_{i=0}^{L}$ is the sampled impulse response of the cascoded transmit, channel and receive filters, assumed slowly time varying, $\{I_{k-i}\}$ is the sequence of transmitted symbols, which are assumed identically distributed independent random variables and $\{n_k\}$ is an additive white noise sequence with Gaussian distribution. At each instant the data takes one of the M possible levels $\{\pm 1, \pm 3, \cdots, \pm(M-1)\}$ with equal probability.

We first consider the problem of estimating a sequence of N transmitted data symbols from a sequence of channel outputs $\mathbf{r}_k = [r_1, r_2, \cdots r_N]'$ for a known channel. There are M^N equi-probable sequences denoted by $\{\mathbf{I}(1), \cdots, \mathbf{I}(M^N)\}$. The ML estimator chooses the most likely sequence $\mathbf{I}_{ML}$ according to

$$\mathbf{I}_{ML} = \arg\max_{\mathbf{I}} f_{\mathbf{r}|\mathbf{I}}(\mathbf{r}|\mathbf{I}), \tag{2}$$

where $f_{\mathbf{r}|\mathbf{I}}(\cdot|\cdot)$ is the conditional probability density function (pdf). Since $\{n_k\}$ are iid random variables, one can write

$$\begin{aligned} f_{\mathbf{r}|\mathbf{I}}(\mathbf{r}|\mathbf{I}) &= \prod_{k=1}^{N} f_{r_k|\mathbf{I}}(r_k|\mathbf{I}_k), \\ &= \prod_{k=1}^{N} f_n(r_k - \mathbf{I}_k'\mathbf{h}), \end{aligned} \tag{3}$$

where $f_n(\cdot)$ is the Gaussian pdf. In principle the maximization in eq. (2) should take place through exhaustive search over the M^N sequences, which can be carried out efficiently using the VA [1].

For the unknown blind equalization problem at hand, we consider the conditional probability of the received sequence $\mathbf{r}_k$ conditioned on both the transmitted sequence and the channel impulse response. Assuming all channel realizations are equally probable, the joint ML estimate for the transmitted data sequence and the channel parameters is given by,

$$(\mathbf{I}_{ML}, \mathbf{h}_{ML}) = \arg\max_{\mathbf{I},\mathbf{h}} f_{\mathbf{r}|\mathbf{I},\mathbf{h}}(\mathbf{r}|\mathbf{I}, \mathbf{h}), \tag{4}$$

wherein the maximization is carried over all the possible channel realizations and transmitted data sequences. Such problem is not trivial since $\mathbf{h}$ is continuous and $\mathbf{I}$ is discrete.

While [3,4] solved eq. (4) by finding the least square estimate of the channel and then using that estimate in the VA, we follow a different approach. The key observation is that

if the channel is discrete, one could interchange the roles of data and channel parameters in the VA branch metric. That is if the data is known one would search a channel trellis for the ML channel parameters and vise versa. Therefore we propose to use two trellises, one for data and the other for channel. Two VAs are used to search two trellises in parallel one for the data and the other for the channel. The output of one would feed the other and the vice versa. This joint maximization would eventually converge to the estimate given in eq. (4).

The channel parameters are approximated by discrete values from the infinite alphabet $\{0, \pm c, \pm 2c, \cdots\}$, where c can be chosen to be arbitrary small. With such a channel alphabet the corresponding channel trellis will have an infinite number of states. However since the channel vector $\mathbf{h}$ does not vary over each signaling interval, as the data vector $\mathbf{I}$, we need not consider all the levels of the possible channel parameters at a given instant. In order to reduce complexity, we propose a simple assignment scheme for the channel trellis. The next channel estimate $\mathbf{h}^{i+1}$, in the proposed scheme is given by

$$\mathbf{h}^{i+1} = h^i \quad \text{for state 0}$$

and

$$\mathbf{h}^{i+1} = h^i \pm c \cdot \mathbf{1}_n \quad \text{for state } n = 1, 2, \cdots,$$

where $\mathbf{1}_n$ ia a vector of length L are either zeros or ones. For the special case when $\mathbf{1}_n = \mathbf{0}$, the degenerate state 0 results. Clearly the number of states[2] does not depend on the parameter c but on the channel memory L.

A smaller number of states can be used if the vector $\mathbf{1}_n$ is restricted to be all zero, except for the element at the n th location to unity. is clear that the above state assignment results in $L + 2$ states. The branches emerging from all the states, except for state 0, have 2 parallel transitions, one corresponding to an increment $(+c)$ and the other to a decrement $(-c)$. There are other state assignment schemes with less than 2^{L+1} states, but the above assignment will result in a simple trellis.

The algorithm will proceed as follows:

- 1. Start with initial channel estimate, $\hat{\mathbf{h}}_{ML} = \mathbf{h}^0$.
- 2. Use the VA to solve for

$$\hat{\mathbf{I}}_{\mathbf{ML}} = \arg\max_{\mathbf{I}} f(\mathbf{r}|\mathbf{I}, \hat{\mathbf{h}}_{ML}),$$

  with the branch metric $\left(r_k - \sum_{i=1}^{L} \hat{h}_i I_{k-i}\right)^2$

- 3. Use the VA to solve for

$$\hat{\mathbf{h}}_{\mathbf{ML}} = \arg\max_{\mathbf{h}} f(\mathbf{r}|\hat{\mathbf{I}}_{ML}, \mathbf{h}_{\mathbf{ML}}),$$

  with the branch metric $\left(r_k - \sum_{i=1}^{L} h_i \hat{I}_{k-i}\right)^2$

[2]The maximum number of states is 2^{L+1}.

- 4. Iterate 2 and 3.

It can be noticed that the algorithm achieves the ML estimate of the channel by adaptively incrementing or decrementing the previous estimate. Using the channel state table above, we change one channel parameter per transition. To improve the speed of convergence one can more states in the channel trellis, that allows one to change two or more parameters a time. This will significantly improve the rate of convergence on the expense of complexity. Thus one can compromise rate of convergence to complexity.

Another parameter that affects the performance is the step parameter c. Choosing c smaller will reduce the rate of convergence, but will improve the error rate. This point is demonstrated in the following simulation example.

III. SIMULATION RESULTS

The algorithm described above was used to equalize the channel (assumed unknown) whose sampled impulse response is given by

$$h(n) = 0.405 \cdot \delta(n) + 0.817 \cdot \delta(n-1) + 0.407 \cdot \delta(n-2),$$

where $\delta(\cdot)$ is the Kronecker delta function. For simplicity binary transmission is assumed, so that $L = 2$, and therefore the channel and data trellises will have 4 states each. These trellises are shown in fig. 1.

The channel was initialized to $\mathbf{h}^0 = (0\ 0\ 0)$. The bit error rate for four different values of c were determined by simulation. The bit error of the ideal MLSE (with known channel) was also determined. These are depicted in fig. 2. With $c = 0.01$, the error rate of the proposed algorithm closely follows that of the ideal MLSE. For larger values of c, the error rate deviates from the ideal case. However for $c = 0.05$ case there is less than 1 dB loss.

The rate of convergence results are not shown due to space limitations. However as expected the rate of convergence for the $c = 0.01$ case was the lowest among the four cases.

IV. CONCLUSION:

In this communication, a new algorithm for blind MLSE was proposed. This algorithm approximates the continuous level channel model by a discrete level one. A channel state assignment scheme was presented, that leads to a simple channel trellis. This together with a data trellis are used to find the joint ML channel and data estimates.

The algorithm offers a good complexity versus performance tradeoff. One can compromise complexity for faster convergence rate and lower error rate. The rate of convergence depends directly on the parameter c. This paper considered using a constant step c, using a variable c would be valuable to acheive a higer rate of convergence. This point is currently pursued by the authors.

References

[1] G.D. Forney, "Maximum-likelihood sequence estimation of digita l sequences in the presence of intersymbol interference," *IEEE Trans. on I nformation Theory*, Vol. IT 18, pp. 363-378, 1972.

[2] GODARD, D. N.: 'Self-recovering equalization and carrier tracking in two-dimensional data communication systems', *IEEE Trans. on Com munications*, 1980, **COM 28**, pp. 1867-1875.

[3] SESHADRI, N.: 'Joint data and channel estimation using fast blind trellis search techniques', *Conference Records of GLOBECOM '90*, pp. 1659-1 663.

[4] GHOSH, M. and WEBER, C.: 'Maximum-likelihood blind equalization' , *Optical Engineering* **31**, pp. 1224-1228.

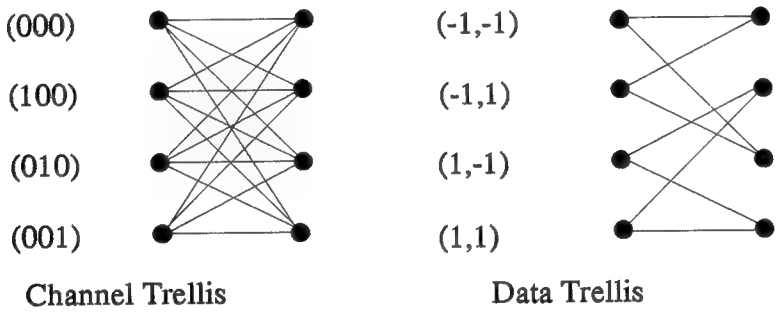


Fig. 1 Channel and date trellis

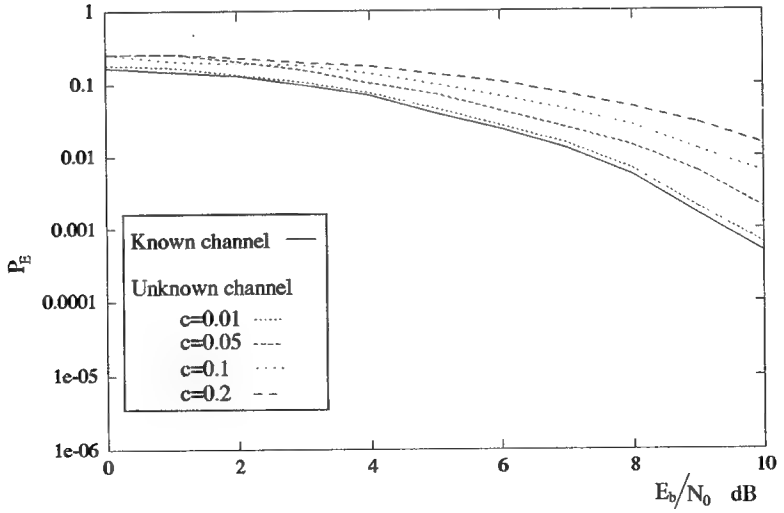


Fig. 2 Bit error rate for different values of c

APPENDIX C: PART IV

REDUCED-COMPLEXITY SEQUENCE ESTIMATIION USING STATE PARTITIONING

by
Raafat Kamel and Yeheskel Bar-Ness

ABSTRACT

A reduced-state sequence estimator for linear dispersive channels is described. The technique is based on partitioning the set of all channel states in a way that defines a trellis with fewer states. The new technique generalizes trellises that were previously obtained by Duel-Hallen and Heegard, and Eyouboğlu and Qureshi. While the approach followed by Duel-Hallen results in powers of two-state trellises, the new technique generates trellises with an arbitrary number of states. Using the state partitioning approach one could also generalize the trellises obtained by Eyouboğlu. Therefore, the approach described in this paper provides a better performance/complexity tradeoff.

I. INTRODUCTION

In high speed digital transmission over bandlimited channels, one of the principal impairments, besides additive Gaussian noise, is intersymbol interference (ISI). Equalization is used to combat ISI. The type of equalization used to mitigate ISI can be divided into two classes. The first, symbol-by-symbol equalization, encompasses linear and decision feedback equalization. The second involves maximum likelihood sequence estimation (MLSE) [1], where the Viterbi algorithm (VA) is used to search a trellis for the desired sequence.

While the first class has low complexity and a high error rate, the second has a lower error rate at the expense of complexity. The complexity of the VA grows exponentially with the length of the channel impulse response. When the impulse response becomes larger, the VA becomes impractical, and methods for complexity reduction are needed.

Research has been directed toward obtaining reduced-complexity equalizers, while maintaining a performance as close as possible to MLSE. To reduce the complexity, a number of authors have proposed incorporating a linear or decision feedback preprocessor so that the MLSE will deal with an equivalent channel having a shorter impulse response [2] [3]. In [2], a linear equalizer was used to shorten the impulse response of the channel, while in [3] a DFE was used to truncate the length of the channel. Such approaches were found to limit the performance of the combined system.

Eyuboğlu and Qureshi [4], and Duel-Hallen and Heegard [5] have proposed sequence estimators which provide a good performance/complexity tradeoff. The reduced-state sequence estimation (RSSE) of [4] is useful for systems utilizing a large signal constellation, while the delayed decision-feedback sequence estimation (DDFSE) of [5], which is a special case of

RSSE, is suitable for channels with a long impulse response.

In [4], trellises with reduced number of states are constructed. These states are formed by combining the states of the maximum likelihood estimator using Ungerboek-like set partitioning principles. An ambiguity arises in the branch metric calculation due to the combination of the different states. A feedback detector is incorporated to resolve this ambiguity.

In [5] the complexity of the VA is reduced by considering a few states of the channel. The ISI due to the rest of the states is estimated using a feedback detector analogous to that of the decision feedback equalizer (DFE). The estimated ISI is then used in the branch metric computation. As in the DFE, error propagation affects the performance of the algorithm. The degradation due to error propagation was found to be less than that of the DFE.

In [6] a sequential algorithm; namely, the stack algorithm, used for decoding convolutional codes, is adapted to ISI channels. A metric suitable for ISI was derived in [6]. The stack algorithm offers an advantage in terms of computational complexity over the VA. This comes with an increase in stack and buffer storage. Reference [7] gives a good survey on different sequential decoding algorithms.

In this paper, a new technique [8] is presented for reducing the complexity of the VA for channels with long memory and an arbitrary size signal constellation. The technique is closely related to [4] [5]. While [4] is suitable for a signal alphabet with large constellations and [5] is suited for channels with long memory, the new approach offers a unified approach to the problem of a large signal constellation with an arbitrarily long channel. It also offers more flexibility in the choice between performance and complexity than the one in [5]. It can generate trellises with any number of states rather than only power of M states as in [5], where M is the alphabet size.

The approach described in this paper is motivated by the work on error propagation for the DFE given in [9]. Error sequences for the DFE can be modeled as a Markov chain, whose number of states is exponential in the number of distinct magnitudes of error and the number of past decisions that influence the current decisions. The complexity of the resulting systems is extremely high. To reduce system complexity Duttweiler, *et al.*, [9] proposed a reduced state machine, wherein error sequences are grouped together in a unique manner. This grouping can also be envisioned as partitioning the set of all possible error sequences. This led to using a state partitioning technique to reduce sequence estimation complexity.

This paper is organized as follows. In section 2, we present the channel model and introduce the partitioning approach for reducing the complexity of the VA for the binary transmission case. In section 3, we describe a general procedure to perform the suggested partitioning. An example is given in section 4. In section 5, we discuss the probability of error for the reduced complexity scheme. The technique is extended to complex non-binary modulation in section 6. Conclusions are given in section 7.

II. CHANNEL MODEL AND THE PROPOSED TECHNIQUE

We will consider channels with finite impulse response, whose discrete time model is given in Figure 1. This model arises in a quadrature amplitude modulation (QAM) system at the output of a sampled, whitened matched filter. Using the D transform, the channel $h(D)$ is

modeled as a finite response filter (FIR), and $n(D)$ is a white Gaussian noise source with a zero mean and variance of σ^2. The data sequence $a(D)$ consists of symbols a_k, which are independent and identically distributed. In this section we will assume binary transmission; extension to complex and non-binary modulation is deferred to section 6. The data symbols a_k take values of ± 1 with equal probability. Referring to Figure 1, the output $y(D)$ is given by

$$y(D) = a(D)h(D) + n(D),$$

where $h(D)$, given by $h(D) = h_0 + h_1 D + \cdots + h_n D^n$, defines the channel impulse response, whose degree, n, is determined by the channel memory. The state of the channel, $s(k)$, at time k is the binary n tuple given by $(a_{k-1}, a_{k-2}, \cdots, a_{k-n})$, the previous input data. At any time k there are 2^n possible states. The set of all channel states is denoted by Ω; *i.e.*,

$$\Omega = \{s_i : s_i \text{ is a state of the system }, i = 0, 1, \cdots, 2^n - 1\}.$$

In the proposed technique different states are combined into a smaller number of super-states. That is, the set Ω is divided into N, S_i, subsets, where N is $2 \leq N \leq 2^n$, such that

1. $\bigcup_{i=0}^{N-1} S_i = \Omega$
2. $S_i \cap S_j = \emptyset$; the empty set, for $i \neq j$ and $0 \leq i, j \leq N-1$
3. The subsets S_i are chosen such that for all $s_n(k) \in S_i$, the corresponding next state $s_n(k+1)$ must belong to one subset.

The first two conditions specify a partition on the set Ω, and hence one could also specify an equivalence relation on Ω. The third condition is a constraint on the partitions that enables a trellis to be defined. It can be shown that not every partition on Ω could be a candidate; only those that result in a trellis are suitable. A procedure is devised for defining partitions that satisfy the third condition. This is detailed in section 3.

At this point one should emphasize the difference between the partitioning considered in this paper and that in [4]. Therefore a brief description of the RSSE is in order.

Consider a signal constellation, $\mathcal{C}$, with M points, the state of the channel $s(k)$ is given by an M-ary n tuple $(a_{k-1}, a_{k-2}, \cdots, a_{k-n})$. The RSSE [4] associates each element a_{k-i}, for $1 \leq i \leq n$, with a set partition $\mathcal{C}(i)$ of the constellation, in which the signal constellation is partitioned into J_i subsets. The number of subsets J_i ranges from 2 to M. In [4] the partition $\mathcal{C}(i)$ is chosen as a finer partition of $\mathcal{C}(i+1)$ and $J_i \geq J_{i+1}$ for $1 \leq i \leq n-1$. Using the constellation partitioning described in [4], the channel state $s(k)$ is represented by a sub-state, defined by

$$t_k = [x_{k-1}(1), x_{k-2}(2), \cdots, x_{k-n}(n)],$$

where $x_{k-i}(i)$ is the subset to which the transmitted symbol a_{k-i} belongs. Now since $x_{k-i}(i)$ can only assume one of J_i possible values, the RSSE will have a total of $\prod_{i=1}^{n} J_i$ states as opposed to M^n original states. When $J_i = M$ for $1 \leq ilel$ and $J_i = 1$ for $l < i \leq n$, RSSE becomes DDFSE.

The trellises generated by the DDFSE represent a subset of those obtained by the RSSE technique. By using the state partitioning approach described in this paper one can obtain more trellis choices. Consider the special case of a binary alphabet, the approaches followed in [4][5] will result in 2^m-state trellises, m ranging from 1 to N. On the other hand the state partitioning approach could generate trellises with an arbitrary number of states not necessarily a power of two.

The branch metric for the MLSE is given by $(y_k - \sum_{i=1}^{n} h_i a_{k-i} - a_k)^2$. Since each state in the reduced trellis is a union of two or more channel states, an ambiguity will result in the branch metric calculation. That is, the branch metric is no longer uniquely determined by the previous/present trellis states' pair. Similar to [4] [5], a feedback mechanism is introduced to resolve this ambiguity. The branch metric associated with the reduced trellis is given by $(y_k - \sum_{i=1}^{l} h_i a_{k-i} - \sum_{l+1}^{n} h_i \hat{a}_{k-i} - a_k)^2$, where $l < m$ is determined by the reduced trellis. The previous state estimate $(\hat{a}_{k-l-1}, \cdots, \hat{a}_{k-n})$ is stored in the path history associated with the present state.

III. THE PARTITIONING PROCEDURE

A partition P of a set Ω is a pairwise disjoint collection of non-empty subsets of Ω, whose union is Ω. It is known that an equivalence relation in Ω defines a partition of Ω, and, conversely, a partition in Ω yields an equivalence relation. Given an equivalence relation R in Ω, let $R(a) \triangleq \{x \in \Omega : aRx\}$ for each $a \in \Omega$. $R(a)$ is known as an equivalence class of R and is a subset of Ω. The collection of subsets, $P = \{R(a) : a \in \Omega\}$, is a partition of Ω. A collection of equivalence relations $\{R_1, R_2, \cdots, R_n\}$ is known as an equivalence sequence iff for all i, j $1 \leq i \leq j \leq n$, and all $x, y \in \Omega$ we have $xR_j y \Longrightarrow xR_i y$. That is, $R_n(x) \subset R_{n-1}(x) \subset \cdots \subset R_1(x) \subset \Omega$.

For the channel model described in the previous section, the states of the channel are given by binary n tuples. Consider the relation R_i given by: $xR_i y$ iff the first i components of the n-tuples x and y are identical, for any states x and $y \in \Omega$. It can be shown that R_i is an equivalence relation. Therefore R_i defines a parition of Ω. It can also be shown that the sequence $\{R_1, R_2, \cdots, R_n\}$ is an equivalence sequence.

Figure 2 shows the partitions associated with different R_i and the corresponding subsets. Label the subsets corresponding to R_i, with binary i tuples. We refer to the partition formed by R_i as level i partition. It should be noted that there are 2^i subsets at level i, each with cardinality 2^{n-i}. The number of subsets in P will determine the number of states of the trellis, that is satisfy the third constraint. We will now show that the equivalence sequence $\{R_1, R_2, \cdots, R_n\}$ can define partitions that will result in a reduced complexity trellis.

In order to show that the partitions associated with the above equivalence sequence satisfy the third constraint, we first present it in a mathematical form.

Define two functions $F_1(x)$ and $F_{-1}(x)$ as the next state of the channel, when the present state is x, and inputs are 1 and -1 respectively. That is, if $x = (a_1, a_2, \cdots, a_n)$, then

$$\begin{aligned} F_1(x) &= (1, a_1, \cdots, a_{n-1}) \\ \text{and} \qquad F_{-1}(x) &= (-1, a_1, \cdots, a_{n-1}). \end{aligned}$$

To meet the third condition, the equivalence relation R must satisfy

$$\begin{aligned} xRy &\implies F_1(x)RF_1(y) \text{ and} \\ &\implies F_{-1}(x)RF_{-1}(y) \quad \text{for all } x \text{ and } y \in \Omega. \end{aligned} \tag{1}$$

When the above statements are satisfied, the functions $F_1(\cdot)$ and $F_{-1}(\cdot)$ are said to be compatible with R. Equation (1) can also be written as,

$$\begin{aligned} \text{for all } x \text{ and } y \in \Omega \quad y \in R(x) &\implies F_1(y) \in R(F_1(x)) \text{ and} \\ &\implies F_{-1}(y) \in R(F_{-1}(x)). \end{aligned} \tag{2}$$

3.1 Trellises with 2^i States

The partition P is formed by considering all the subsets from the ith level; in other words all the equivalent classes of the equivalent relation R_i. We can now show that such partitions satisfy the third constraint, *i.e.*, equations (1) or (2). Consider the equivalence relation R_i; the set $\{F_1(y) : y \in R_i(x)\}$ for some $x \in \Omega$ is the set of all channel states that have the first $i+1$ components identical. The first element is 1 and the consecutive i elements are identical, since the previous state $y \in R_i(x)$. Therefore, one can write:

$$\begin{aligned} \{F_1(y) : y \in R_i(x)\} &= R_{i+1}(F_1(x)) \\ &\subset R_i(F_1(x)). \end{aligned}$$

The second step follows from the definition of an equivalence sequence. Therefore, it can be concluded that

$$xR_iy \implies F_1(x)R_{i+1}F_1(y) \tag{3}$$

$$\text{Therefore} \quad \implies F_1(x)R_iF_1(y) \qquad \text{for all } x \text{ and } y \in \Omega. \tag{4}$$

A similar argument holds for $F_{-1}(\cdot)$; that is,

$$\begin{aligned} xR_iy &\implies F_{-1}(x)R_{i+1}F_{-1}(y) \text{ and} \\ &\implies F_{-1}(x)R_iF_{-1}(y) \quad \text{for all } x \text{ and } y \in \Omega. \end{aligned} \tag{5}$$

Comparing with equation (1), we conclude that the equivalence relation R_i defines a partition which satisfies the third condition and would result in a trellis. Since the equivalence relation R_i partitions the set Ω into 2^i subsets, the resulting trellis will have 2^i states. It should be mentioned that these trellises are the same as those found by Duel-Hallen, *et al.* [5] On the other hand, as will be shown in the next section, by using the state partitioning technique, one can find trellises with any number of states, rather than only power of 2.

3.2 Trellises with Number of States not 2^i

In this case the partition P is formed by considering subsets from adjacent levels, instead one level as in the previous section. This can be done by considering some level i subsets and the remaining subsets are replaced by their corresponding subsets from level $i+1$. This will satisfy the first two conditions given in section 2. To satisfy the third condition one has to meet two requirements:

1. For all x and $y \in \Omega$

$$xR_iy \implies F_1(x)R_iF_1(y) \quad \text{if} \quad R_i(F_1(x)) \in P \text{ or} \tag{6}$$
$$\implies F_1(x)R_{i+1}F_1(y) \quad \text{otherwise} \tag{7}$$

2. For all x and $y \in \Omega$

$$xR_{i+1}y \implies F_1(x)R_{i+1}F_1(y) \text{ if } R_{i+1}(F_1(x)) \in P \text{ or} \tag{8}$$
$$\implies F_1(x)R_iF_1(y) \quad \text{otherwise .} \tag{9}$$

$F_{-1}(x)$ has to satisfy similar requirements.

The requirements given in (6) and (8) follow from equation (4). Condition (7) also follows from equation (4). Now since

$$F_1(x)R_{i+1}F_1(y) \implies F_1(x)R_iF_1(y) \quad \text{from definition of equivalence sequence,}$$

and

$$xR_{i+1}y \implies F_1(x)R_{i+1}F_1(y) \quad \text{from equation (4) ,}$$

it follows that

$$xR_{i+1}y \implies F_1(x)R_iF_1(y),$$

and (9) is satisfied. A similar argument holds for $F_{-1}(x)$. Therefore, the partition P formed by considering subsets from two adjacent levels results in a trellis. The number of states of such a trellis will not be a power of 2.

It is worth mentioning that partitions formed by subsets from non-adjacent levels will not form a trellis since (7) will not be satisfied. In fact,

$$xR_iy \implies F_1(x)R_jF_1(y) \text{ and}$$
$$\implies F_{-1}(x)R_jF_{-1}(y) \quad \text{for all } x \text{ and } y \in \Omega$$

is true only for $j = i+1$.

IV. AN EXAMPLE

Consider the channel given by

$$h(D) = h_0 + h_1D + h_2D^2 + h_3D^3. \tag{10}$$

The above channel has a memory of $n = 3$; therefore, the state can be represented by binary 3 tuples $x = (x_1, x_2, x_3)$. One can use up to three levels of partitioning, or the equivalence sequence $\{R_1, R_2, R_3\}$. Using the notation in section 3, Table 1 gives different partitioning schemes. Consider the 5-state trellis given in Table 1. The set of all states Ω is partitioned by considering subsets $R_2(01)$, $R_2(1,0)$ and $R_2(1,1)$ from level 2 and instead of subset $R_2(0,0)$ we consider its subsets $R_3(000)$ and $R_3(001)$ from level 3. Note that there are $\begin{pmatrix} 4 \\ 3 \end{pmatrix} = 4$ possible trellises with 5 states. The last entry in Table 1 is the full complexity case of 8 states. The branch metric depends on the originating node of a given branch. If the originating node of a branch corresponds to a subset from level l ($l < n$), then the branch metric γ_k is given by

$$\gamma_k = (y_k - \sum_{i=0}^{l} h_i a_{k-i} - \sum_{i=l+1}^{n} h_i \hat{a}_{k-i})^2. \tag{11}$$

The trellises for the partitioning schemes considered in Table 1 are given in Figure 3.

V. PROBABILITY OF ERROR

In this section we will investigate the error performance of the partitioning schemes developed in section 3. It was noted in [4] [5] [10] that the effect of the error propagation is minimal for moderate to high SNR. Therefore, we will assume in our derivation that the effect of error propagation is negligible. In the sequel we will consider trellises with powers of two states, from which other trellises will be derived.

5.1 Trellises with 2^i States

These trellises are the same as those derived in [5], and hence the analysis given in [5] applies here. Nevertheless, we will relate the probability of error to different partitioning levels. This will be vital for the analysis of trellises with an arbitrary number of states.

Consider the trellis formed by the subsets from level i. As noted earlier, the resulting trellises will have 2^i states, which are represented by the binary i tuple. Consider the channel at time k, the channel state is $s(k)$; $s_i(k) = (a_{k-1}, a_{k-2}, \cdots, a_{k-i})$, let the corresponding state estimated by the VA be denoted by $\hat{s}_i(k)$, $\hat{s}_i(k) = (\hat{a}_{k-1}, \hat{a}_{k-2}, \cdots, \hat{a}_{k-i})$. Following Forney's approach [1], an error event $\mathcal{E}$ is said to occur between $k = k_1$ and $k = k_2$, if $\hat{s}_i(k_1) = s_i(k_1)$, $\hat{s}_i(k_2) = s_i(k_2)$ and $\hat{s}_i(k) \neq s_i(k)$ for all k, $k_1 < k < k_2$. Since $\hat{s}_i(k) = s_i(k)$ for $k = k_1, k_2$, it follows that

$$(\hat{a}_{k_1-1}, \hat{a}_{k_1-2}, \cdots, \hat{a}_{k_1-i}) = (a_{k_1-1}, a_{k_1-2}, \cdots, a_{k_1-i})$$

and

$$(\hat{a}_{k_2-1}, \hat{a}_{k_2-2}, \cdots, \hat{a}_{k_2-i}) = (a_{k_2-1}, a_{k_2-2}, \cdots, a_{k_2-i}).$$

Now define the input error sequence associated with the event $\mathcal{E}$, by

$$\mathbf{e} \triangleq \{e_{k_1}, e_{k_1+1}, \cdots, e_{k_2-i-1}\},$$

where $e_k \triangleq a_k - \hat{a}_k$. The Euclidean distance $d_i^2(\mathcal{E})$ of the error event is given by

$$d_i^2(\mathcal{E}) = \sum_{k=k_1}^{k_2} \left(\sum_{j=0}^{\min(k-k_1,i)} h_j e_{k-j} \right)^2 . \tag{12}$$

In the case of binary transmission, the probability of error is upper bounded by [1]

$$P_e \leq \sum_{d \in D} Q\left(\frac{d_i}{2\sigma}\right) \sum_{\mathcal{E} \in E_{d_i}} w(\mathbf{e}) 2^{-w(\mathbf{e})},$$

where E_{d_i} is the set of all error events having a Euclidean distance of d_i^2 and D is the set of square roots of Euclidean distances attained by error events. The factor $w(\mathbf{e})$ is the number of bit errors a given error event entails, and $Q(\cdot)$ is given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy.$$

For moderate to high SNR the upper bound of the probability of error is dominated by events attaining the minimum distance, *i.e.*,

$$P_e \leq K_i Q\left(\frac{d_{i_{min}}}{2\sigma}\right), \tag{13}$$

where K_i is given by

$$K_i = \sum_{\mathcal{E} \in E_{d_{i_{min}}}} w(\mathbf{e}) 2^{-w(\mathbf{e})}. \tag{14}$$

Note that we used the subscript i throughout to emphasize the dependence of terms like $d_{i_{min}}$ and K_i on the level of partitioning i. Therefore, to evaluate the upper bound on the probability of error for a given level, one has to determine $d_{i_{min}}$ and K_i.

At lower SNR one can get better bounds by considering the stack algorithm given in [10]. However, with the stack algorithm one has to first find the error state diagram. The complexity of such a diagram becomes intractable for channels with a long impulse response. Therefore, we will only consider events with minimum distances.

5.2 Trellises with Number of States not 2^i

We showed in section 3 how to form trellises by considering subsets from adjacent levels rather than from one level. Examples of these trellises were given in section 4.

To find an upper bound on the probability of error for these trellises, we consider without loss of generality the trellis formed by subsets from levels i and $i+1$. Furthermore assume that the trellises are formed by considering p subsets from level i and q from level $i+1$ in such a way that the third constraint given in section 2 is satisfied. At moderate to high

SNR, using the total probability theorem, the probability of error of such a scheme can be upper bounded by

$$P_e \leq \frac{p}{2^i} K_i Q\left(\frac{d_{i_{min}}}{2\sigma}\right) + \frac{q}{2^{i+1}} K_{i+1} Q\left(\frac{d_{i+1_{min}}}{2\sigma}\right). \tag{15}$$

It can be easily shown that for $q = 0$ the upper bound for level i results, while for $p = 0$ that of level $i + 1$ results. Note that in equation (15) the first term dominates asymptotically, since $d_{i_{min}} < d_{i+1_{min}}$. That is, at high SNR the first term in equation (15) is more dominant than the second. Therefore, the performance of such a trellis would be the same as that of level i at sufficiently high SNR. However, at moderate SNR, the performance of these trellises is better than those with 2^i states, *i.e.*, trellises formed by considering subsets from level i only. This point is demonstrated in the following example. The improvement in performance becomes more pronounced for longer channels.

5.3 Simulation and Upper Bounds

As an example, consider the channel whose impulse response is given by

$$h(n) = 0.7107{\cdot}\delta(n)+0.1421{\cdot}\delta(n-1)+0.2132{\cdot}\delta(n-2)+0.1421{\cdot}\delta(n-3)+0.6396{\cdot}\delta(n-4). \tag{16}$$

The above channel has a memory of $n = 4$; therefore, the states can be represented by binary 4 tuples $x = (x_1, x_2, x_3, x_4)$. One can use up to four levels of partitioning, or the equivalence sequence $\{R_1, R_2, R_3, R_4\}$. We will consider the trellises formed by the equivalence relations R_3 and R_4. Using the same notation as above, Table 2 gives some schemes we considered in the simulation and computation of the upper bounds.

For level 4, the error sequences that have minimum distance are $\pm(2, 0, 0, 0, 0)$, and for level 3 these sequences are $\pm(2, 0, 0, 0)$. The simulation results and the upper bounds are shown Figure 4. Referring to Figure 4, there is less than a 2dB loss when considering the 8-state instead of the 16-state trellis. As expected with 12- and 14-state trellises the probability of error is better than the 8-state. The amount of improvement decreases as SNR increases. At high SNR the improvement becomes insignificant; that is, the 12- and 14-state trellises approach the 8-state performance.

VI. NON-BINARY MODULATION

In this section the concept of state partitioning is extended to include non-binary modulation schemes. The state $s(k)$ is represented by an M-ary n tuple $(a_{k-1}, a_{k-2}, \cdots, a_{k-n})$, where a_k is an M-ary symbol given by

$$a_k = \begin{cases} -(M-1)\cdot d, \cdots, -d, d, \cdots (M-1)\cdot d & \text{for M-PAM} \\ x_k + jy_k, \text{ where } x_k \text{ and } y_k \text{ take values} & \\ -(\frac{M}{2}-1)\cdot d, \cdots, -d, d, \cdots, (\frac{M}{2}-1)\cdot d & \text{for M-QAM} \\ d\exp(\frac{2\pi}{M}i), i = 0, 1, \cdots, M-1 & \text{for MPSK,} \end{cases}$$

where d is a function of the transmitted power. The number of states in this case is given by M^n. Thus the set of channel states Ω has a cardinality of M^n.

Following a similar procedure as in section 3, the equivalence sequence $\{R_1, \cdots, R_n\}$ is used, where R_i is again defined as xR_iy iff the first i components of the n-tuples x and y are identical for all x and $y \in \Omega$. The partitioning tree in this case will have M branches emerging from each node. Reduced complexity trellises can be constructed by considering subsets or union of subsets from different levels. This is best illustrated by the following example.

Consider the transmission of a 4-QAM source (Figure 5) through a channel whose impulse response is given by

$$h(D) = h_0 + h_1 D + h_2 D^2 + h_3 D^3.$$

In this case the channel has $4^3 = 64$ states. Since the memory of the channel is 3, the partition tree can have up to 3 levels. The partitioning tree is shown in Figure 6. Trellises with different numbers of states can be constructed using the subsets shown in Figure 6. Table 3 lists some of these trellises. Table 3 also indicates an RSSE-type trellis with the same number of states, if they exist. Those given in the table were arbitrarily chosen. Since the channel has a memory of three, for the RSSE we need to specify three parameters (J_1, J_2, J_3), which are given in Table 3. As with the state partitioning technique there could be more than an RSSE-type trellis for a given number of states, an arbitrary one is given in the table. For illustration purposes consider the 8-state case. In Figure 7a we present the trellis obtained using the state partitioning approach, which is the same the RSSE trellis [4] which is depicted in Figure 7b. For the trellis shown in Figure 7b, we used $J_1 = 4$, $J_2 = 2$ and $J_3 = 1$. It can be seen from Table 3 that the state partitioning approach results in a wider range of trellises than with the RSSE. As a result it offers a better performance/complexity tradeoff compared to the RSSE. No attempt has been made to compare the distance properties of these trellises, since we just intended to show the existence of a wider range of trellises, some of which could not be produced by the RSSE [4] approach.

VII. CONCLUSION

In this paper we introduced a new approach to reduce the complexity of the VA. This approach is based on partitioning the set of channel states. It also a offers good complexity versus performance tradeoff. It was shown that the trellises obtained in [5] and [4] are special cases of those described in this paper.

The state partitioning approach results in trellises with an arbitrary number of states which are not necessarily powers of 2 states, as in [5] for a binary signal set. Depending on the length of the channel and the operating SNR, trellises with non-power of 2 states can attain a lower probability of error than the powers of 2. However, at high SNR the improvement of the former over the latter is insignificant. The technique can also be extended to include complex and non-binary signal sets.

References

[1] G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. on Information Theory*, Vol. IT 18, pp. 363-378, 1972.

[2] S.U.H. Qureshi and E.E. Newhall, "An adaptive receiver for data transmission over time-dispersive channels," *IEEE Trans. on Information Theory*, Vol. IT 19, pp. 448-457, 1973.

[3] W. U. Lee and F. S. Hill, "A maximum-likelihood sequence estimator with decision-feedback equalization," *IEEE Trans. on Communications*, Vol COM 25, pp. 971-979, 1977.

[4] M.Y. Eyuboğlu and S.U.H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. on Communications*, Vol. COM 36, pp. 13-20, 1988.

[5] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. on Communications*, Vol. COM 37, pp. 428-436, 1989.

[6] F. Xiong, A. Zerik and E. Shwedyk, "Sequential sequence estimation for channels with intersymbol interference of finite or infinite length," *IEEE Trans. on Communications*, Vol. COM 38, pp. 795-804, 1990.

[7] J. B. Anderson and S. Mohan, "Sequential coding algorithms: a survey and cost analysis," *IEEE Trans. on Communications*, Vol. COM 32, pp. 169-176, 1984.

[8] R. E. Kamel and Y. Bar-Ness, "Reduced state estimation of digital sequences in dispersive channels using state partitioning" *Electronics Letters*, pp. 14-16, Jan. 6, 1994.

[9] D.L. Duttweiler, J.L. Mazo and D.G. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *IEEE Trans. on Information Theory*, Vol. IT 20, pp. 490-497, 1974.

[10] W. H. Sheen and G. L. Stüber, "Error probability for reduced-state sequence estimation," *IEEE Journal on Selected Areas in Communications*, Vol. JSAC 10, 571-578.

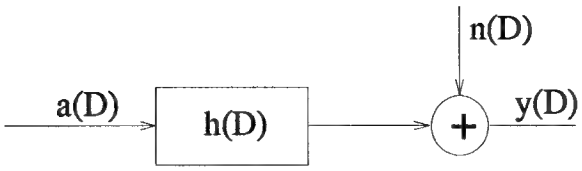


Figure 1: The discrete channel model

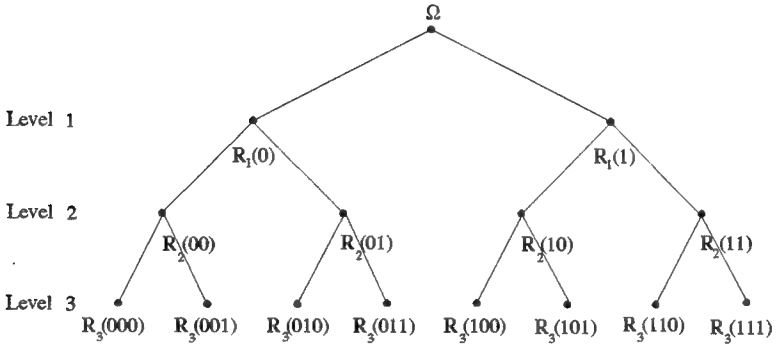


Figure 2: The partitioning tree for a binary source

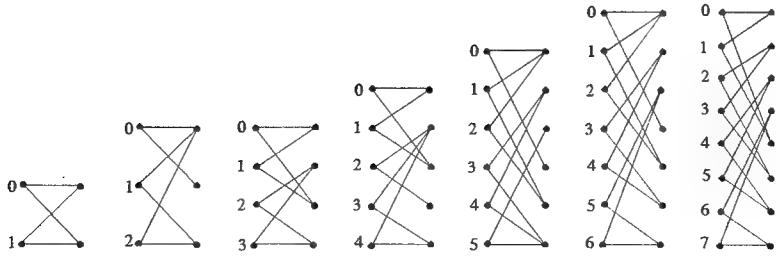


Figure 3: Trellises for the schemes given in Table 1

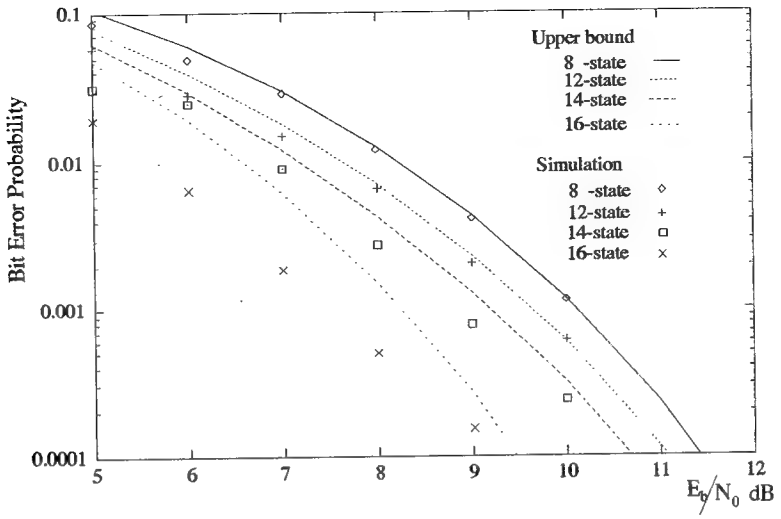


Figure 4: Probability of error for different partitioning schemes

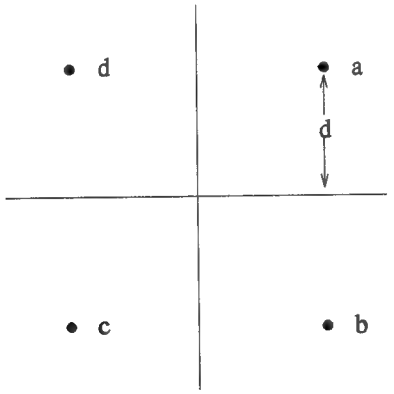


Figure 5: Constellation for 4-QAM.

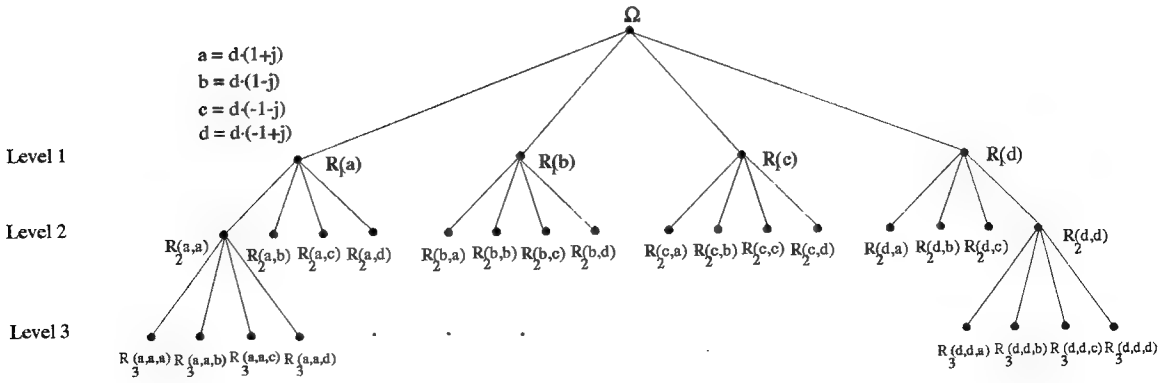


Figure 6: Partitioning tree for a channel with a memory of 3 and a 4-QAM source.

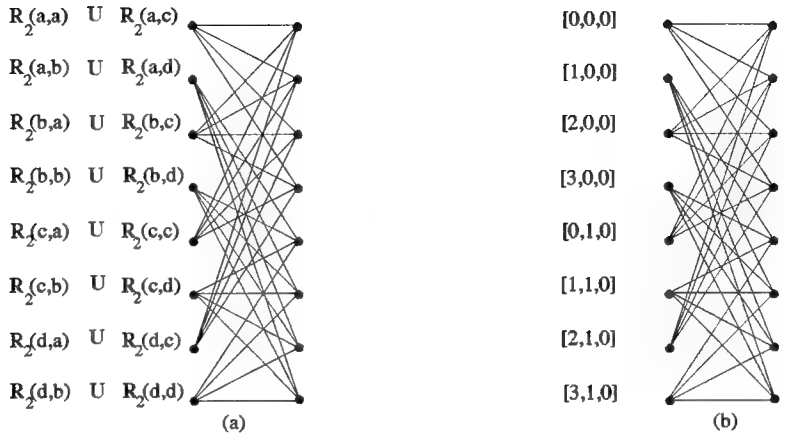


Figure 7: (a) The 8-state trellis obtained using the state partitioning approach. (b) The 8-state trellis obtained using RSSE [4].

Number of States	Partitions
2	$\{R_1(0), R_1(1)\}$
3	$\{R_1(0), R_2(10), R_2(11)\}$
4	$\{R_2(00), R_2(01), R_2(10), R_2(11)\}$
5	$\{R_3(000), R_3(001), R_2(01)$ $R_2(10), R_2(11)\}$
6	$\{R_3(000), R_3(001), R_2(01)$ $R_2(10), R_3(110), R_3(111)\}$
7	$\{R_3(000), R_3(001), R_2(01)$ $R_3(100), R_3(101), R_3(110), R_3(111)\}$
8	$\{R_3(000), R_3(001), R_3(010), R_3(011)$ $R_3(100), R_3(101), R_3(110), R_3(111)\}$

Table 1: Different partitions for a channel with a memory of $n = 3$.

Number of States	Partitions
8	$\{R_3(000), R_3(001), R_3(010), R_3(011)$ $R_3(100), R_3(101), R_3(110), R_3(111)\}$
12	$\{R_3(000), R_3(001)$ $R_4(0100), R_4(0101), R_4(0110), R_4(0111)$ $R_4(1000), R_4(1001), R_4(1010), R_4(1011)$ $R_3(110), R_3(111)\}$
14	$\{R_3(000), R_4(0010), R_4(0011)$ $R_4(0100), R_4(0101), R_4(0110), R_4(0111)$ $R_4(1000), R_4(1001), R_4(1010), R_4(1011)$ $R_4(1100), R_4(1101), R_3(111)\}$
16	$\{R_4(0000), R_4(0001), R_4(0010), R_4(0011)$ $R_4(0100), R_4(0101), R_4(0110), R_4(0111)$ $R_4(1000), R_4(1001), R_4(1010), R_4(1011)$ $R_4(1100), R_4(1101), R_4(1110), R_4(1111)\}$

Table 2: Selected partitions for a channel with a memory of $n = 4$

# of States	Partitions	RSSE (J_1, J_2, J_3)
2	$\{R_1(a) \cup R_1(c),\ R_1(b) \cup R_1(d)\}$	(2,1,1)
3	$\{R_1(a), R_1(b) \cup R_1(d), R_1(c)\}$	(3,1,1)
4	$\{R_1(a), R_1(b),\ R_1(c),\ R_1(d)\}$	(2,2,1)
5	$\{R_2(a,a), R_2(a,b), R(a,c), R(a,d)$ $R_1(b) \cup R_1(c) \cup R_1(d)\}$	does not exist for RSSE
6	$\{R_2(a,a), R_2(a,b), R(a,c), R(a,d)$ $R_1(b) \cup R_1(d), R_1(c)\}$	(3,2,1)
7	$\{R_2(a,a), R_2(a,b), R(a,c), R(a,d)$ $R_1(b), R_1(c), R_1(d)\}$	does not exist for RSSE
8	$\{R_2(a,a) \cup R_2(a,c),\ R_2(a,b) \cup R_2(a,d)$ $R_2(b,a) \cup R_2(b,c),\ R_2(b,b) \cup R_2(b,d)$ $R_2(c,a) \cup R_2(c,c),\ R_2(c,b) \cup R_2(c,d)$ $R_2(d,a) \cup R_2(d,c),\ R_2(d,b) \cup R_2(d,d)\}$	(4,2,1)
10	$\{R_2(a,a), R_2(a,b),\ R_2(a,c), R_2(a,d)$ $R_2(c,b), R_2(c,b),\ R_2(c,c), R_2(c,d)$ $R_1(b), R_1(d)\}$	does not exist for RSSE
12	$\{R_2(a,a), R_2(a,b),\ R_2(a,c), R_2(a,d)$ $R_2(b,a) \cup R_2(b,b),\ R_2(b,c) \cup R_2(b,d)$ $R_2(c,a), R_2(c,b),\ R_2(c,c), R_2(c,d)$ $R_2(d,a) \cup R_2(d,b),\ R_2(d,c) \cup R_2(d,d)\}$	(4,3,1)
14	$\{R_2(a,a), R_2(a,b),\ R_2(a,c), R_2(a,d)$ $R_2(b,a), R_2(b,b),\ R_2(b,c), R_2(b,d)$ $R_2(c,a), R_2(c,b),\ R_2(c,c), R_2(c,d)$ $R_2(d,a) \cup R_2(d,c),\ R_2(d,b) \cup R_2(d,d)\}$	does not exist for RSSE
16	$\{R_2(a,a), R_2(a,b),\ R_2(a,c), R_2(a,d)$ $R_2(b,a), R_2(b,b),\ R_2(b,c), R_2(b,d)$ $R_2(c,a), R_2(c,b),\ R_2(c,c), R_2(c,d)$ $R_2(d,a), R_2(d,b),\ R_2(d,c), R_2(d,d)\}$	(4,4,1)
⋮		
64	$\{R_3(a,a,a), \cdots, R_3(d,d,d)\}$	(4,4,4)

Table 3: Different partitions for a channel with a memory of n=3 and a 4-QAM source.

*U.S. GOVERNMENT PRINTING OFFICE: 1995-610-126-50210

Rome Laboratory

Customer Satisfaction Survey

RL-TR-_____

Please complete this survey, and mail to RL/IMPS, 26 Electronic Pky, Griffiss AFB NY 13441-4514. Your assessment and feedback regarding this technical report will allow Rome Laboratory to have a vehicle to continuously improve our methods of research, publication, and customer satisfaction. Your assistance is greatly appreciated.
Thank You

_____

_____

Organization Name:_____(Optional)

Organization POC: _____(Optional)

Address:_____

1. On a scale of 1 to 5 how would you rate the technology developed under this research?

5-Extremely Useful 1-Not Useful/Wasteful

Rating_____

Please use the space below to comment on your rating. Please suggest improvements. Use the back of this sheet if necessary.

2. Do any specific areas of the report stand out as exceptional?

Yes___ No_____

If yes, please identify the area(s), and comment on what aspects make them "stand out."

3. Do any specific areas of the report stand out as inferior?

Yes___ No___

If yes, please identify the area(s), and comment on what aspects make them "stand out."

4. Please utilize the space below to comment on any other aspects of the report. Comments on both technical content and reporting format are desired.

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.